EVALUATION OF A DENSITOMETRIC METHOD FOR DETECTING NULL GENES OF THE FOURTH COMPONENT OF COMPLEMENT, C4

CENTRE FOR NEWFOUNDLAND STUDIES

TOTAL OF 10 PAGES ONLY
MAY BE XEROXED

(Without Author's Permission)

DEIRDRE-ANN BOISSEAU, B.Sc.

EVALUATION OF A DENSITOMETRIC METHOD FOR DETECTING NULL GENES OF THE FOURTH COMPONENT OF COMPLEMENT, C4

by

© Deirdre-Ann Boisseau, B.Sc.

A thesis submitted to the School of Graduate
Studies in partial fulfillment of the
requirements for the degree of
Master of Science

Faculty of Medicine
Memorial Univerity of Newfoundland
April 1989

St. John's Newfoundland

To my *late* grandmother

Albertine Boisseau

ABSTRACT

The fourth component of complement (C4) is encoded by tandem MHC-linked genes, C4A and C4B. C4A and C4B have multiple alleles, including nulls, at each locus. Individuals who are homozygous null for C4A or C4B are identified by the absence of C4A or C4B in serum. The complete absence of C4A is strongly associated with systemic lupus erythematosus (SLE). Although the C4A or C4B heterozygous-null genotype can only be assigned with certainty when informative family data are available, many investigators use relative intensity of C4A/C4B bands after serum electrophoresis to assign null alleles. These assignments are based on unproven assumptions regarding gene-dose and relative C4A/B expression, and pose a serious problem in population studies that attempt to show associations between null C4 alleles and disease, as for example in SLE. In this study the usefulness of C4A:C4B ratio measured by densitometry of typing bands as a method for detecting null alleles was examined.

C4A:C4B densitometric ratios were determined for 108 individuals genotyped by family studies. Discriminant analysis was used to compare genotype derived from family data with genotype predicted from A:B ratio. Mean ratios for various C4 genotypes and total C4 concentrations by single radial immunodiffusion were also evaluated. C4A:C4B ratios were then determined for six populations of unrelated individuals including SLE and rheumatoid arthritis (RA) patients and control subjects. Observed frequencies for genotypes predicted by A:B ratio were compared to expected frequencies, assuming Hardy-Weinberg equilibrium. MHC data were evaluated for individuals mis-classified by densitometry, and those with very high and very low ratios.

The results of this study showed that the overall accuracy of the densitometric method was approximately 80%. The main error in the method was the false identification of subjects as heterozygous null for C4A. The overall mis-classification trend toward the heterozygous null C4A group suggested that many individuals may have more C4B than C4A. Using densitometric measures of C4A:C4B ratio to estimate

population frequencies of carriers of null C4 genes tended to overestimate the frequency of carriers of null C4A. The overestimation was pronounced in populations in which the actual frequency of the null C4A gene was low. The assumptions upon which the densitometric method is based may not be valid because the MHC haplotype HLA-B17, C4A6C4B1, HLA-DR7 may carry a C4A6 that is expressed at low levels, and the haplotype HLA-B44, C4A3C4BQ0 may carry two copies of C4A3 or a single A3 that is expressed at high levels.

In summary, the ratio of C4A to C4B in an individual's serum is an inaccurate predictor of his/her genotype.

ACKNOWLEDGEMENTS

I express thanks to my supervisor Dr. V. Skanes and my committee members, Drs. B. Larsen and D. McKay. Dr. Y. LeGal deserves special thanks for allowing me to use his densitometer, the piece of equipment which was essential to this project. Thanks also to Dr. G. Skanes for making his SYSTAT statistical package available for my use. I also thank Mrs. K. Ross for her humour and patience in introducing me to the joys of word-processing used to prepare this thesis.

Fellowship support was provided by the School of Graduate Studies of Memorial University of Newfoundland and from the Faculty of Medicine as bursary support.

I express my deepest thanks to my family and to Jim for the understanding and moral support they have shown during the course of my graduate work.

Lastly, I dedicate this thesis to my late grandmother, Albertine, a lady of great character and stamina, whose memory inspired me throughout the long hours spent on this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	=	analysis of variance
BF	=	Factor B
°C	=	Celsius
C2	=	the second component of complement
C4	=	the fourth component of complement
CIE	=	crossed immunoelectrophoresis
cm	=	centimeter
CPseB	=	carboxypeptidase-B
CV	=	co-efficient of variation
EDTA	=	ethylene-diamine-tetraacetic acid
HLA	=	human leucocyte antigens
IFE	=	immunofixation gel electrophoresis
kb	=	kilobase
mA	=	milliampere
mg	=	milligram
MHC	=	major histocompatibility complex
mM	=	millimolar
mm	=	millimeter
NAse	=	neuraminidase
nm	=	nanometer
OD	=	optical density
PAGE	=	polyacrylamide gel electrophoresis
PBS	=	phosphate-buffered saline
PMSF	=	phenyl-methyl-sulfonamide
RA	=	rheumatoid arthritis
RR	=	relative risk
SD	=	standard deviation
SDS	=	sodium-dodecyl-sulfate
SLE	=	systemic lupus erythematosus
SRID	=	single radial immunodiffusion
Tris	=	tris(hydroxymethyl)aminomethane
µl	=	microliter
v	=	volts
W	=	watts

Chapter 1

INTRODUCTION

1.1. Historical background

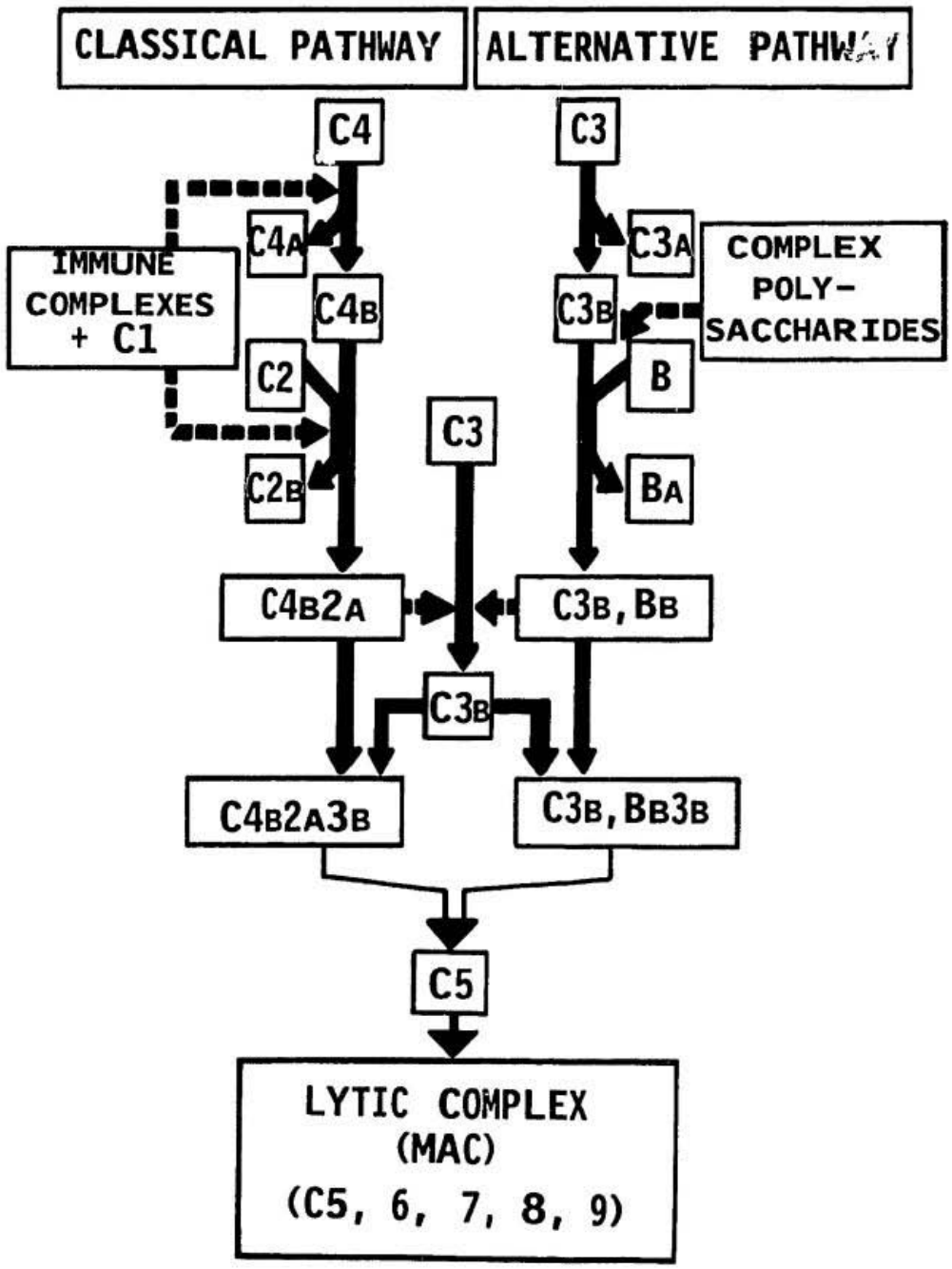
The complement system was first recognized to be important in host defence against infection in 1894, by Jules Bordet. Much of the seminal work in immunology by Ehrlich, Bordet, and Metchnikoff focused on the role of antibody and complement in immunity to infection (Ross, p.xi, 1986). As early as 1926 a physiological role for the fourth component of complement (C4) in the human immune system had been shown (Gordon, Whitehead and Wormall, 1926). However, it was not until the advent of modern methods of protein separation and characterization in the 1960's, that the nature and mechanism of the complement system began to be uncovered.

1.2. The complement system

The complement system as known, is composed of twenty distinct plasma proteins, several membrane proteins, and nine different membrane receptors. The plasma proteins normally circulate in an inactive form. A major function of the complement system is to destroy both foreign microorganisms and immune complexes. Thus, the complement system is activated by polysaccharides such as those found in yeast and bacterial cell walls, and complexes of foreign antigen and specific antibody. There are two known pathways of activation, the classical pathway and the alternative pathway (Figure 1.1). The classical pathway is activated primarily by antigen-antibody aggregates; while the alternative pathway is usually activated by the complex polysaccharides of cell walls. Activation of the complement system results in a series of

Figure 1.1 **The classical and alternative pathways of complement activation.** C1, C2 and C4 are the early acting components of the classical pathway. Factor B is the alternative pathway equivalent of C2. Both pathways lead to the formation of C3 convertase and subsequent assembly of the late acting components, C5-C9. These terminal components form the membrane attack complex (MAC), which effects lysis of target cell membranes.

NOTE: Most of the components of both complement pathways are normally represented by lower case letters (e.g. C4b2a) as distinguished from upper case letters used for alleles (e.g. C4A). However, only upper case letters were available in the font size used to prepare this figure. The larger letters thus represent capitals, the smaller letters represent lower case letters.


CLASSICAL PATHWAY
ALTERNATIVE PATHWAY
C4
C3
C4A
C3A
IMMUNE COMPLEXES + C1
C4B
C3B
COMPLEX POLY-SACCHARIDES
C2
B
C3
C2B
BA
C4B2A
C3B, BB
C3B
C4B2A3B
C3B, BB3B
C5
LYTIC COMPLEX (MAC)
(C5, 6, 7, 8, 9)

biochemical reactions in which inactive precursor molecules are enzymatically cleaved to form active components. This results in a series of responses that eliminate the foreign molecule or cell from the tissues or circulation.

Of the complement proteins, the fourth component of complement, C4, is essential to the classical pathway of complement activation (Reid and Porter, 1981), (Figure 1.1). Beside its role in forming C3 convertase, C4 forms a covalent bond with the target immune complex. This binding is important for assembly of later components at the activation site and for immune clearance. When native C4 is activated by proteolytic cleavage, an internal thioester bond found near the centre of the C4 α-chain (C4d region) is exposed. This unusual structure releases a reactive acyl group which forms a covalent bond with the $-NH_2$ or -OH groups of antigen-antibody complexes (Campbell, Gagnon and Porter, 1981; Harrison, Thomas and Tack, 1981).

1.3. The genetics of C4

Interest in the genetics of C4 came from several observations: cases of complete and partial C4 deficiencies that resulted in immune complex diseases (Hauptmann, Grooshans and Reid, 1974); localization of the C4 genes within the major histocompatibility complex (MHC) (Rittner *et al.* 1975); and evidence of the extreme polymorphism of C4 (Awdeh and Alper, 1980).

1.3.1. The major histocompatibility complex (MHC)

The MHC is a gene complex which was first recognized because the products of its genes include cell surface glycoproteins which are responsible for tissue graft rejection among members of the same species. In humans these products are called the human leukocyte antigens (HLA). Located on the short arm of chromosome 6, the MHC is divided into three regions termed Class I, Class II and Class III. The cell surface glycoproteins which are involved in self-nonself discrimination, are encoded in the

regions of the MHC called Class I and Class II. The Class I antigens are referred to as HLA-A, -B, and -C, and these are the classical transplantation antigens present on all nucleated cells of the body. The Class II antigens have a much more restricted distribution, occurring primarily on cells involved in the antigen presentation phase of an immune response. The genes for the Class II antigens are located in the D-region of the MHC and encode the proteins HLA-DP, HLA-DQ and HLA-DR. The Class III region contains the genes which code for the fourth (C4) and second (C2) components of the classical complement pathway, Factor B (BF) of the alternative complement pathway, 21-hydroxylase A and 21-hydroxylase B (for a review of the structure of the MHC, see Campbell *et al.* 1988).

Evidence for linkage of C4 to HLA came initially from studies of the murine H-2 complex. In the mid-1970's various groups showed that the serum substance (Ss) protein which Shreffler and Owen (1963) had mapped to the murine MHC (H-2), was structurally (Curman *et al.* 1975; Hansen and Shreffler, 1975; Meo, Krasteff and Shreffler, 1975) and functionally (Carroll and Capra, 1978) similar to human C4. As electrophoretic variants and deficiencies of C4 were studied, a more precise mapping of the C4 genes became possible. In 1976 Teisberg and his co-workers demonstrated structural polymorphism of C4. They concluded that the bands they observed, after electrophoretic separation of serum C4, represented the expression of two codominant alleles at one locus--and that this was the structural locus for C4. They determined that this genetic locus was linked to HLA by studying a number of informative HLA-B/C4 meioses. They found no recombinations--strong evidence that the two loci were closely linked.

Independent of the electrophoretic studies, the Chido (Middleton *et al.* 1974) and Rodgers (Giles *et al.* 1976) 'blood group' antigens were shown to be controlled by two genes which were linked to HLA. In 1978, O'Neill, Yang and Dupont used family studies to show that C4 polymorphism was controlled by not one, but two distinct, although closely linked, loci (O'Neill *et al.* 1978a). The same year, they found that the

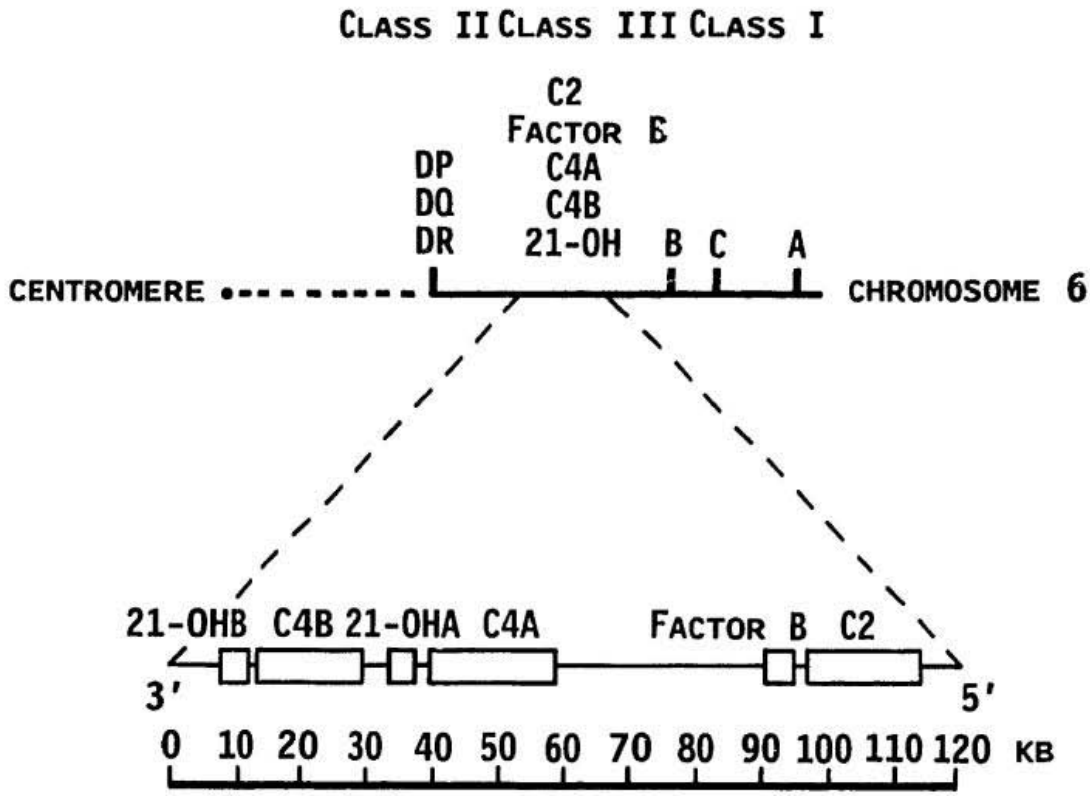
Chido and Rodgers erythrocyte antigens were antigenic determinants of human C4 (O'Neill *et al.* 1978b).

In addition to the two C4 loci, genetic studies had already positioned two other complement genes, the second component, C2 (Fu *et al.* 1974; Meo *et al.* 1977) and Factor B or BF (Alper, Boenisch and Watson, 1972; Allen, 1974) in the Class III region. In 1977 Dupont *et al.* positioned the genes for steroid 21-hydroxylase (21-OH) to the Class III region.

By the mid-1980's, the application of molecular genetics techniques had produced a more refined map of the MHC. Several groups established that the C4 genes were positioned between HLA-B and HLA-DR (Weitkamp and Lamm, 1982; Carroll *et al.* 1984a; Robinson *et al.* 1985). The order of the complement genes relative to one another was determined by Carroll and his colleagues in 1984 (Carroll *et al.* 1984b). However, the orientation and distance of the Class III genes relative to Class I and Class II genes had not been established. In 1987, using the powerful new technique of pulsed-field gel electrophoresis, two groups answered this question, producing the most refined map of the MHC to date (Dunham *et al.* 1987; Carroll *et al.* 1987a), (Figure 1.2). The order of the complement genes centromeric of HLA-B is C2, BF, C4A, 21-OHA, C4B, 21-OHB.

The C2 and BF genes, which are separated by less than 1kb, lie about 30kb from the C4A gene (Campbell, Bentley and Morley, 1984). The C4A locus is separated from the C4B locus by approximately 10kb. The 21-OH genes are located less than 2kb 3' to each C4 gene (White *et al.* 1985).

The number of C4 genes varies among individuals, with one two or three loci being found per chromosome. In Caucasian populations, most individuals (50-65%) inherit four functional genes; 30-38% have three functional genes, (i.e. are heterozygous for one null/non-expressed allele); and 5-10% have only two functional genes (Hauptmann, Tappeiner and Schifferli, 1988). Individuals with only one C4 gene are very rare (<1%).



(Adapted from Carroll, 1987c)

Figure 1.2 **Molecular map of the MHC region containing the genes for C4A, 21-hydroxylase A (21-OHA), C4B, 21-hydroxylase B (21-OHB), C2 and Factor B (BF).**

1.4. Structure of C4

The human liver is the major site of plasma C4 synthesis (Nagura *et al.* 1985). Macrophages are the chief extrahepatic site of C4 synthesis (Colten and Dowten, 1986).

C4 is a large glycoprotein. In plasma, C4 consists of three polypeptide chains, α (95kD), β (75kD) and γ (30kD), linked by disulfide bridges (Reid and Porter, 1981), (Figure 1.3). Mouse studies have shown that C4 is synthesized as a single chain precursor of about 200kD, with the polypeptides in the order β, α, and γ (Sim and Sim, 1981; Shreffler *et al.* 1984).

1.5. C4 polymorphism

C4 is highly polymorphic. Polymorphism of C4 was first studied by Rosenfeld and co-workers (1969) using crossed immunoelectrophoresis of fresh EDTA-plasma.

1.5.1. Detection of C4 polymorphism

Evidence for genetic polymorphism of human C4 has come from the techniques summarized in Table 1.1. The standard techniques for C4A and C4B variant classification are: prolonged alkaline agarose-gel electrophoresis of EDTA-plasma or serum followed by immunofixation (IFE), (Alper and Johnson, 1969) and alkaline agarose-gel electrophoresis followed by hemolytic gel overlay (Awdeh and Alper, 1980; Mauff *et al.* 1983). IFE separates variants based on their electrophoretic mobilities. Additionally, functional differences of C4 isotypes can be identified using the hemolytic gel overlay. In the former technique, serum or plasma is electrophoresed in agarose and the different C4 protein bands are detected by in-gel fixation using polyclonal antibody specific for C4. The antigen-antibody complexes are thus trapped in the gel, and all protein except for the complex can be washed away. The bands are then stained with protein stain. In the latter technique, the detection system is a hemolytic gel overlay containing sensitized sheep erythrocytes and C4-deficient serum. This method identifies

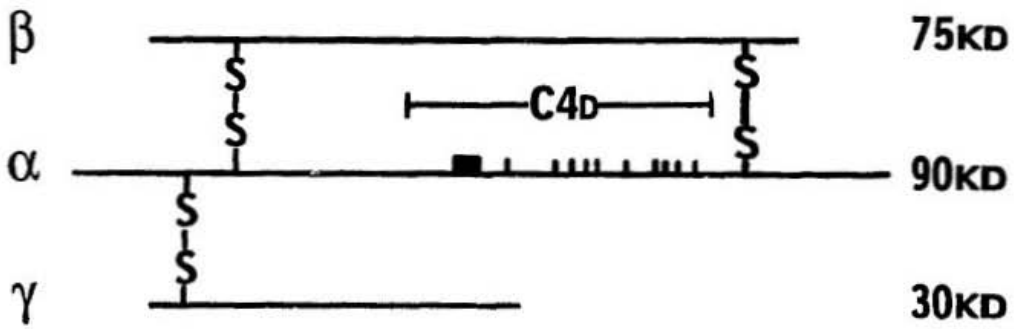


Figure 1.3 **Schematic representation of plasma C4.** The three polypeptide chains α, β, and γ, are linked by disulfide (S-S) bridges. The amino (NH_2) ends are on the left of the figure, the carboxy (COOH) ends are on the right. Vertical lines in the α-chain represent isotypic and allotypic differences localized to the C4d region. The solid block represents the internal thioester involved in covalent binding reactions.

TABLE 1.1

Techniques for detecting polymorphism of C4 *

	METHOD	INFORMATION
1.	Immunofixation agarose-gel electrophoresis	> 30 allelic banding patterns
2.	Hemolytic detection on agarose-gel electrophoresis	C4B allotypes > C4A allotypes
3.	Agglutination with anti-Rg and anti-Ch sera	Rodgers (Rg) and Chido (Ch) antigen determinations
4.	Crossed immunoelectrophoresis	- Null alleles by relative quantitation of C4A and C4B
5.	C4 α- and C4 β-chain determination on SDS-gel electrophoresis	Null alleles by relative quantitation of: - C4A-α and C4B-α bands - C4 β_H and C4 β_L bands
6.	Immunoblotting of C4 after SDS gel electrophoresis with Rg and Ch antisera	- Localization of Rg/Ch epitopes to C4 polypeptide chains
7.	Immunoblotting of C4 after agarose-gel electrophoresis with C4B monoclonal antibody	- Localization of C4B epitope to particular C4 allotypic variants
8.	DNA typing with C4 cDNA	- Number of C4 genes - DNA polymorphism - Subdivision of allotypes

* Adapted from Rittner and Mauff, 1984a

C4A and C4B variants based on their different hemolytic efficiencies, because C4B is approximately four times more hemolytically active than C4A (Dodds, Law and Porter, 1985), (Table 1.2). Hemolytic gel overlay is often combined with IFE to differentiate between C4A and C4B variants whose bands overlap. Another C4-phenotyping method, crossed immunoelectrophoresis (CIE), is quite often used to detect individuals having one null C4 allele (Awdeh, Raum and Alper, 1979). This method is based on the gene-dose effect which proposes that all C4A and C4B genes are expressed equally and that individuals do not carry hidden C4A or C4B gene duplications. C4 proteins are separated by an initial electrophoresis, then subjected to a second electrophoresis which drives the proteins into antibody incorporated in the second gel. The resulting antigen-antibody precipitates form as distinct peaks and the area under each peak is proportional to the concentration of the individual protein. Thus, according to the gene-dose hypothesis, an individual with two C4A alleles and one C4B allele should have a C4A peak approximately twice the height of the C4B peak.

Over the years several methods of C4 phenotyping by conventional agarose gel immunofixation electrophoresis were reported (Teisberg *et al.* 1977; O'Neill *et al.* 1978a; Mauff, Bender and Fischer, 1978). None of these methods differed substantially, and in 1980, Awdeh and Alper introduced a simple technical maneuver which revealed the extensive polymorphism at both C4 loci. The modification they introduced was neuraminidase (NAse) treatment of whole plasma or serum, prior to electrophoresis. Neuraminidase removes sialic acid residues from the glycosylated C4, with the result that the products of the two C4 loci migrate with much less overlap than they do in the native state. The latest modification to IFE is the pre-treatment of serum or plasma with carboxypeptidase-B (CPseB), prior to NAse treatment (Sim and Cross, 1986). CPseB treatment resolves the triplet pattern seen with NAse alone, into a single sharp band for each allotype. To visualize the complexity of the electrophoretic separation of C4 when NAse is used alone, consider that most people inherit four different C4 genes. This implies an electrophoretic pattern could contain twelve (or more) bands! CPseB

TABLE 1.2

Comparison of human C4A and C4B proteins: functional, structural and serological differences [a]

		C4A	C4B
1.	Electrophoretic mobility		
	(a) Agarose gel	Fast (acidic)	Slow (basic)
	(b) SDS-PAGE (α chain)	Mw 96 000	Mw 94 000
2.	Thioester reactivity		
	(a) Hemolytic activity	Lower	Higher
	(b) Relative covalent binding affinities		
	(i) Amino group	Higher	Lower
	(ii) Hydroxyl group	Lower	Higher
3.	Antigenic determinants *	Rodgers (Rg: 1,2)	Chido (Ch: 1,2,3,4,5,6)

* with exceptions such as C4A1 and C4B5

[a] Adapted from Yu, Campbell and Porter, 1988

is a porcine pancreatic enzyme which cleaves C-terminal arginine and lysine residues (Folk *et al.* 1960). From sequence analysis of C4 cDNA (Belt, Carroll and Porter, 1984), the C-terminus of the C4α-chain would be expected to have four C-terminal arginine residues, and the C-terminus of the C4β-chain would be expected to have the sequence arg-lys-lys-arg. Sim and Cross (1986) hypothesize that the difference in migration of each of the (three) C4 bands seen with NAse treatment alone, represents small charge differences due to residual carbohydrate determinants, all of which would be removed by the clearing of the four terminal amino acids. They suggest therefore, that after removal of the C-terminal basic amino acids by CPseB, the charge on all the C4α and all the C4β-chains is the same. As a result, the minor bands are converted into the (single) most anodally migrating band.

Currently, 35 alleles have been assigned to the combined C4A and C4B loci, including non-expressed or null alleles (termed Q0, for "quantity zero") (Mauff *et al.* 1983). In 1983, in an effort to standardize results being obtained from C4 reference laboratories, a standard immunofixation electrophoresis technique was agreed upon, as was a common nomenclature of C4 allotypes (Mauff *et al.* 1983). Although many of the C4 variants are rare, common variants occur (in Caucasians) with a frequency of at least 0.7% (Table 1.3). The 35 alleles identified probably represent a miminum, since the most common alleles, C4A3 and C4B1, are being subdivided on the basis of serological determinants (Giles, 1984); DNA restriction-fragment length polymorphism (RFLP) (Whitehead *et al.* 1983); and nucleotide sequencing (Belt *et al.* 1985). Clearly, as the common allotypes are subdivided by these techniques, the nomenclature based on protein typing will have to be reviewed.

1.6. C4 nomenclature

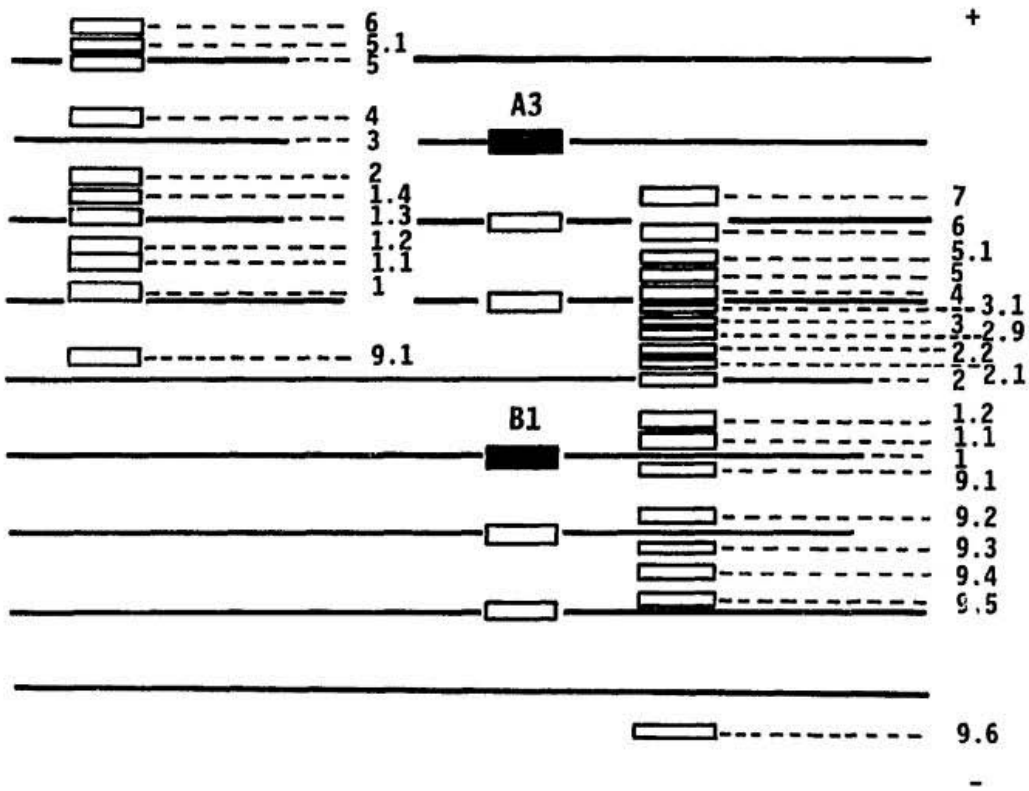
The protein with the more acidic (anodal) migration is called C4A, whereas the protein with the more basic (cathodal) migration is called C4B (Mauff *et al.* 1983). If samples are not treated with carboxypeptidase-B, each allotype pattern usually consists of at least one (anodal) major band and, in most cases, two (cathodal) minor bands (Figure 1.4). The common allotypes are classified by a numerical system from A1 through A6,

TABLE 1.3

Gene frequencies of common C4A and C4B variants in Caucasians

C4 allele	Frequency in Caucasians (%) [a]
A2	6.4
A3	70.2
A4	6.1
A6	3.8
AQ0	12.5
B1	71.4
B2	9.1
B3	3.2
B4	0.7
B5	0.7
BQ0	13.7

[a] Baur *et al.* 1984



(Adapted from Mauff *et al.* 1983)

Figure 1.4 **Schematic representation of C4A and C4B variants after electrophoretic separation of neuraminidase-treated serum or plasma.** Electrophoretic separation of serum or plasma, treated with only neuraminidase, produces a C4-banding pattern as shown above for the common variants C4A3C4B1 (centre). C4A is the more anodal (+) protein, C4B is the more cathodal (-) protein. Usually each allotype pattern consists of one leading major band (solid block in figure) and two minor bands. The allotypes are classified numerically. C4A variants are shown on the left, C4B variants are shown on the right. With the exception of C4A3B1, the variants are indicated by the position of only the most anodal band (+).

and from B1 through B7 (Mauff *et al.* 1983). The standared nomenclature for C4 allotypes (Mauff *et al.* 1983), uses an asterisk to identify alleles, e.g. C4A*3, C4B*1. For simplicity however, asterisks are not used in this text.

1.7. Properties of C4A and C4B

There are two classes or isotypes of C4, C4A and C4B. Although C4A and C4B are highly homologous proteins structurally, they show marked differences in several respects (Table 1.2). The two isotypic forms of C4 differ dramatically in efficiency of covalent binding to proteins and carbohydrates (Isenman and Young, 1984; Law, Dodds and Porter, 1984). C4A exhibits higher covalent binding affinity to amino groups or peptide antigens, with the result that it binds more efficiently to immune complexes than C4B. C4A is also more efficient at inhibiting immune precipitation than C4B (Schifferli *et al.* 1986). C4B binds preferentially to hydroxyl groups or carbohydrate antigens. This observation may explain the (four-fold) higher hemolytic activity of C4B in the conventional hemolytic assay which uses sensitized sheep erythrocytes, whose surface is rich in carbohydrate antigens.

Comparison of C4A and C4B cDNA sequences indicates 14 nucleotide differences, of which 12 are clustered in the C4d region of the α-chain, causing 9 amino acid substitutions (Belt *et al.* 1984), (Figure 1.3). Four of these substitutions are responsible for C4A-B class differences, while the rest contribute to allelic variation (Yu *et al.* 1988).

While it has been shown that isotypic differences do affect function, so far, with the exception of the C4A6 variant, allotypic variation affecting function has not been observed. C4A6 has been found to code for a hemolytically inactive product, usually associated with HLA-B17 (Bw57) (Dodds *et al.* 1985).

The amino acid substitutions in the C4d region are also responsible for the antigenic determinants Rg and Ch (Yu *et al.* 1986). C4A variants are usually Rg-positive, whereas C4B variants are Ch-positive. Two of the less common C4 alleles, C4A1 and

C4B5, have shown reversed antigenicities (Yu *et al.* 1986). Therefore, while they express their own class specific properties, the A1 allele is Ch positive, and the B5 allele is Rg positive. It should be noted that this chimeric nature is not found in all of the C4A1 and C4B5 allotypes studied so far.

In summary, the limited variation in amino acid sequence, concentrated in the C4d region of the α-chain, is responsible for the distinct electrophoretic mobilities, chemical reactivities (hemolytic activity), covalent binding affinities, and antigenic determinants (Ch and Rg), of C4A and C4B variants.

1.8. Heterogeneity of C4 gene size

C4 genes are large (Carroll *et al.* 1984a), and can differ in size (Yu *et al.* 1986; Schneider *et al.* 1986; Palsdottir *et al.* 1987a). C4 genes are either 22kb or 16kb long, and are referred to as long and short genes, respectively (Palsdottir *et al.* 1987a). All of the C4A genes analyzed to date are of the long form (Palsdottir *et al.* 1987a). Approximately half of the C4B genes are of the short form, the remainder are long (Schneider *et al.* 1986; Palsdottir *et al.* 1987a). The difference in size is due to the presence or absence of a 6-7 kb intron near the 5' end of the gene. Most C4A3B1 haplotypes contain B1-long (Palsdottir *et al.* 1987a). Other C4B allotypes which are coded for by short C4 genes include: B1 (in C4A6,B1), B2 (in C4A3,B2), and B3 (in C4A3,B3) (Palsdottir *et al.* 1987a).

Both the coding and non-coding regions of C4 are highly conserved. All the C4A and C4B alleles are 99% identical. While Class I and Class II proteins are highly homologous, comparison of their derived amino acid sequences with C4 shows no evidence of homology (Carroll *et al.* 1985a). This result suggests that the C4 genes are not located in the Class III region because of a common precursor. Porter (1983) suggested that the two C4 genes are nearly identical either because the gene duplication which likely gave rise to the two C4 genes is a recent event, or there is a functional reason. He pointed out that the products of both loci must be able to perform the same type of functions

in the classical pathway, but may do so with different efficiencies because of allelic differences. For example, he suggested that immune aggregate dissolution may vary with C4 polymorphism. This has particular implications for autoimmune diseases such as systemic lupus erythematosus (SLE) (p. 25).

1.9. C4 null alleles

One revelation from the early studies on C4 genetics was the high frequency of silent or null alleles at either locus (C4AQ0, 5-15%; C4BQ0, 10-20%) (O'Neill *et al.* 1978a; Schendel, O'Neill and Wank, 1984; Partanen and Koskimies, 1986).

In this thesis, null alleles will be identified using the following terms:

(1) Heterozygous null - The occurrence of a single null allele, either C4A or C4B, on one haplotype, but not on the other.

(2) Homozygous null - The occurrence of null alleles at both A or both B loci, i.e. the complete absence of one locus product.

(3) Hemizygous null - The occurrence of a C4AQ0 and a C4BQ0 allele together in an individual, but on different haplotypes.

Null alleles are defined by the absence of C4A or C4B protein in the plasma (O'Neill *et al.* 1978a; Awdeh and Alper, 1980). Clearly, this definition is based on studies of families that contain homozygous null individuals. While individuals who are homozygous null for either C4A or C4B have no detectible product for the null loci, individuals with one null allele at a locus (heterozygotes), will have a detectible product for the other allele at that locus.

1.9.1. Heterogeneity of C4 null alleles

Recent DNA analyses show that null alleles are a heterogeneous category, consisting of deleted genes and non-deleted, yet unexpressed, genes (Carroll *et al.* 1985a; Schneider *et al.* 1986; Yu and Campbell, 1987; Uring-Lambert *et al.* 1987). Unequal

cross-over during meiosis would explain duplications on some haplotypes and deletions on others (Carroll and Alper, 1987b). In other situations where the gene is present but unexpressed, the molecular basis of the null allele has not been defined. Non-expression may be due to defects in transcription, translation, or regulatory genes (Hauptmann *et al.* 1987); the C4 gene may not have been duplicated on all chromosomes (Carroll *et al.* 1985a); or it may be due to a hypothesized gene-conversion mechanism (Yu and Campbell, 1987; Palsdottir *et al.* 1987b). Current knowledge of deleted and non-deleted null genes is discussed below.

1.9.2. Molecular bases of deleted null C4 genes

DNA analyses have shown that about 60% of the null alleles are due to deletions of 28kb of DNA (Schneider *et al.* 1986; Carroll, 1987c). The deletions include either the 21-OHA or 21-OHB gene. The three common deletions include C4A and 21-OHA; C4B and 21-OHA; or C4B and 21-OHB genes (Carroll *et al.* 1985a,b; Schneider *et al.* 1986; Carroll, 1987c). Deletions of C4A and 21-OHB together have not been observed, nor have deletions including both C4A and C4B genes.

Gene conversion has been proposed as a model to explain non-deleted null alleles (Palsdottir *et al.* 1987b; Yu and Campbell 1987). The model suggests that some null alleles result because one locus has been converted to express the product of the other locus. Gene conversion could result in "homo-expression" or "iso-expression". Homo-expression is the expression of two identical allotypes at each C4 locus. From RFLP analysis, Yu and Campbell (1987) have inferred that the HLA-B44, C4A3, C4BQ0, DR6, haplotype is probably C4A3,A3 rather than C4A3,BQ0. They propose that homo-expression would account for the high frequency of C4 null alleles in the population. In individuals with homo-expressed C4A or C4B, null alleles would be assigned by plasma typing because the two C4 loci express phenotypically indistinguishable allotypes, rather than because of deletion of a C4 gene.

Iso-expression is the expression of C4 alleles of the same isotype but different allotype (e.g. C4A3,A2,BQO) on one chromosome. This hypothesis would explain the findings of Schneider *et al.* (1986) that all the haplotypes with non-deleted C4B null genes contain genes which are the same size as C4A genes (i.e. C4B-long).

Muir *et al.* (1984) and Wisnieski *et al.* (1987) have reported another kind of C4 deficiency which is caused by hyposynthesis of C4, transmitted as an autosomal dominant trait with no MHC linkage.

1.10. Complete C4 deficiency

Neither deletions of C4A and 21-OHB together nor deletions including C4A and C4B genes have been observed. Thus, complete C4 deficiency due to homozygosity for the C4AQOC4BQO haplotype (Awdeh, Ochs and Alper, 1981; Sjöholm, Kjellman, and Low, 1985) is not likely the result of deletion of the C4 genes. Hauptmann *et al.* (1987) have suggested that the total absence of C4 protein is probably related to other mechanisms, such as the presence of regulatory genes, or defects at the level of transcription or translation of the C4 genes. Such mechanisms have been shown in individuals who have C2 null genes which are not deleted (Cole *et al.* 1985).

1.11. Association of C4 null haplotypes with specific HLA-types

Most deleted C4A null genes are linked to HLA-B8, DR3 (Carroll *et al.* 1985a; Schneider *et al.* 1986; Palsdottir *et al.* 1987a; Uring-Lambert *et al.* 1987). However, non-deleted C4A null genes show no linkage to specific HLA-types. In contrast to non-deleted C4A genes, most non-deleted C4B null genes segregate with HLA-B44, and about half of these with DR4. All deleted C4B null genes which include deletion of 21-OHA, occur with C4A3, and various HLA-types (Schneider *et al.* 1986).

1.12. Identifying C4 null alleles

Individuals who are homozygous null at either the C4A or C4B locus may be easily identified. Using the conventional typing method of immunofixation electrophoresis, one of the C4 bands will be absent. In contrast, the identification of individuals who are heterozygous null is difficult. It is a major problem of disease association studies that attempt to show an association between C4 null A or B, and a given disease. In the absence of family data, various means have been used by different investigators to detect null alleles in heterozygous individuals. These generally involve inference from protein typing, based on the assumption that individuals with one A or B null allele will have lower levels of C4A or C4B than individuals with neither. Thus, reduced hemolytic activity of A or B, lower than normal levels of A or B observed visually or densitometrically after IFE, PAGE, or by CIE, have been used to assign null alleles (Hauptmann *et al.* 1984; Schifferli *et al.* 1986; Maillet *et al.* 1987). In addition, RFLP analysis has been used to define null alleles at the DNA level. However, the problem in determining true gene number by DNA analysis, comes from null alleles which are not caused by gene deletion (i.e. not expressed).

1.13. C4 levels in relation to C4 null alleles

Presently, there is no clear proof whether a direct relationship exists between the number of functional C4 genes an individual has, and the level of C4 in his/her plasma. One difficulty is that there is a wide range of C4 concentrations in both healthy and diseased individuals. The normal range for serum C4 in healthy Caucasians is 0.20-0.50 mg/ml. However, values range from as low as 0.06 mg/ml to as high as 0.90 mg/ml (Nor-Partigen Table of Reference Values, Hoechst Canada Inc., 1987).

If a simple gene-dose effect were in play and no other regulatory mechanisms existed, one would expect for four, three, two, and one expressed C4 gene(s), relative concentrations of 100%, 75%, 50% and 25%. Awdeh, Ochs and Alper (1981) reported

mean values of 113%, 89%, 56% and 39% for four, three, two and one gene(s), respectively. Several groups have shown that there is, at best, a very rough correlation between mean C4 levels in serum and the number of C4 genes expressed in any individual (Olaisen, Teisberg and Jonassen, 1980; Awdeh, Ochs and Alper, 1981; Sjöholm *et al.* 1985). In these studies there was a high standard deviation for serum values corresponding to each gene group. Thus the means were not all significantly different, and there was considerable overlap between groups. The results suggest that MHC-linked genetic factors are not the only determinants of plasma C4 concentrations.

Correlating gene number with total serum C4 levels is difficult for several reasons. One is the assumption that C4A and C4B genes are expressed equally. Another is that gene number according to family pedigree is accurate, i.e. that there are no hidden (homo) duplications. Heterozygous null individuals pose a particular problem, because the molecular basis of some null genes, and hence the true gene number, is not yet defined. Also an individual's total serum C4 level can be influenced by various factors such as hormones (Averill and Bernal, 1984). What is required is a method which determines individual C4A and C4B levels. Combined with an accurate measure of gene number, the expression of C4A and C4B relative to the number of C4A and C4B genes, could be reliably determined.

1.14. Complement and disease

Autoimmune diseases are characterized by the production of auto-antibodies. That is, "self" somehow becomes recognized as "foreign". When complement plays a part in these illnesses, it generally appears to be performing its normal function i.e. opsonizing targets and inducing inflammation. Complement-mediated disease results when inappropriate targets are attacked, as when inflammation damages normal tissues.

Some examples of autoimmune diseases with which complement alleles are associated are: insulin dependent diabetes mellitus (IDDM), systemic lupus erythematosus

(SLE), rheumatoid arthritis (RA), and Graves' disease. While complement alleles are associated with these diseases, complement polymorphism has not been implicated directly. Therefore, the complement association may be secondary to HLA, and may serve to mark high risk haplotypes.

1.14.1. Complement and SLE

The etiology of SLE is unknown, however multiple factors including genetic, endocrine and environmental ones seem to be involved (Agnello, 1987).

1.14.1.1. Association of null C4 alleles with SLE

Several studies have shown an increased frequency of homozygous C4A null, but not homozygous C4B null, in Caucasians with SLE (Dawkins *et al.* 1983; Fielder *et al.* 1983; Reveille *et al.* 1985; Howard *et al.* 1986; Kemp *et al.* 1987). The average reported incidence of homozygous C4AQ0 in these studies is 12% in SLE patients, as compared to 1% in normal controls. The homozygous null state is associated with a relative risk (RR) of 24 for SLE (Christiansen *et al.* 1983; Fielder *et al.* 1983; Howard *et al.* 1986). However, this state accounts for only a minority of SLE patients (Batchelor *et al.* 1987). Although the risk for SLE in individuals with just one C4 null allele appears to be much less than for the homozygous null state, it is still significant (Christiansen *et al.* 1983; Fielder *et al.* 1983; Howard *et al.* 1986).

Protein studies have reported the frequency of heterozygous null C4A in SLE as 42-60%, compared with 17-19% in healthy individuals (Fielder *et al.* 1983; Reveille *et al.* 1985; Howard *et al.* 1986). Both Fielder and Reveille and their respective colleagues established heterozygosity for C4A using family haplotype data, which was compared with a visual assessment of C4 protein after IFE and/or CIE. Howard's work was based on C4 phenotypes of randomly selected patients and controls. DNA analysis, which distinguishes heterozygous null individuals on the basis of whether or not the null gene

is deleted, confirmed that C4A null is increased in SLE (Kemp *et al.* 1987). Kemp *et al.* reported that in 34.5% of SLE patients the C4A gene was deleted on one chromosome, compared with 12.5% of controls ($p<0.05$). Using the same techniques, Goldstein reported figures of 25% for SLE compared with 16% of controls. Five of their 22 patients were black, and presently, no clear association has been shown between C4AQ0 and SLE in blacks (Goldstein *et al.* 1988). Clearly, the nature of the C4AQ0 association with SLE will have to be re-evaluated, as the molecular bases of null alleles become more fully understood.

The most common extended haplotype among Caucasian SLE patients is HLA-B8, C4AQ0B1, DR3 (Fielder *et al.* 1983; Reveille *et al.* 1985). However, because of the linkage disequilibrium between HLA-B8, DR3 and C4AQ0 (Awdeh *et al.* 1983), it is difficult to assess the relative contributions of the HLA-DR3 and C4AQ0 alleles to disease susceptibility.

Perhaps the most compelling evidence for a primary association of C4A null genes with SLE comes from cases of complete C4-deficiency. This condition is exceedingly rare, with only seventeen cases reported world-wide (Hauptmann *et al.* 1987). However, all of these individuals have SLE or lupus-like illness (Hauptmann *et al.* 1974; Schaller *et al.* 1977; Tappeiner *et al.* 1978, 1982; Ballow *et al.* 1979; Minta *et al.* 1981; Kjellman *et al.* 1982; Mascart-Lemone *et al.* 1983; Klein *et al.* 1984; Dumas *et al.* 1986).

Kemp *et al.* (1987) and Goldstein *et al.* (1988) have shown that the genetic basis for C4 null alleles in SLE is heterogeneous. Up to one third of patients have a large C4A, 21-OHA deletion affecting one chromosome, and in Caucasians this deletion occurs almost exclusively with HLA-B8, DR3. Among non-HLA-B8, DR3 SLE patients heterozygous null for C4A, most showed no deletion of C4A.

Most SLE patients homozygous null for C4A have the gene deleted from both chromosomes. Some show a C4A gene deletion on only one chromosome, suggesting the presence on the other chromosome of one non-deleted, non-functional gene.

1.14.2. Mechanism of C4 involvement in SLE

The mechanism by which total C4 deficiency increases the risk of SLE may be related to a decrease in the solubilization and clearance of immune complexes, in which the classical pathway of complement activation plays a crucial role (Takahashi *et al.* 1978; Hauptmann *et al.* 1986). The deficiency of an early complement pathway component, such as C4, increases the likelihood that antigen-antibody complexes will precipitate, leading to tissue damage and inflammation. C4A appears to bind more efficiently to protein-antigen complexes than does C4B, which suggests that C4A may be the C4 component most important in the clearance of immune complexes in SLE.

1.15. Complement and rheumatoid arthritis

Rheumatoid arthritis (RA) is a rheumatic disease whose principle manifestation is inflammation in and around the joints.
(For a discussion of complement and the rheumatic diseases, see Ross, pp. 204-210, 1986). Rheumatoid Factors (RFs) are the predominant autoantibodies present in RA (Carson, 1981), and they activate complement by the classical pathway. It has been hypothesized that the initiation and perpetuation of RA is due to the accumulation of immune complexes within the joint. These immune complexes fix complement, resulting in an inflammatory response. Moreover, there is evidence that the factor which inhibits solubilization of these complexes may be RF (Naama, Mitchell and Whaley, 1983).

DR4 and specifically certain Dw subtypes of DR4 (Dw4 and Dw14), are strongly associated with RA (Stastny 1976, 1978; Nepom, Seyfried and Holbeck, 1986). Three common DR4-containing haplotypes identified in RA are HLA-B15, DR4; HLA-B40,

DR4; and HLA-B44, DR4 (O'Neill *et al.* 1982; Dyer *et al.* 1984). O'Neill *et al.* (1982) also identified a rare C4 variant, C4B2.9, which is increased in adult RA patients (RR 8.5). They found that this variant is strongly associated with HLA-A2, B15(w62), Cw3, DR4, C4A3. DR1 is also significantly increased in RA (Sacha and Kirwan, 1986; Gregersen, Silver and Winchester, 1987). Gregersen *et al.* (1986) found that particular sequences in the DRβ_1 chains appear to determine susceptibility to RA.

1.15.1. C4 null alleles and RA

The occurrence of particular haplotypes associated with null C4 alleles in RA (e.g. HLA-B44, DR4 with C4BQ0), and the observation that null C4 alleles were associated with other autoimmune diseases, suggested that null C4 alleles might be associated with susceptibility to RA. However, the occurrence of C4 null alleles, particularly C4B null, in RA seems to be associated more with particular systemic complications which arise in some RA patients, than with the RA population as a whole (Thomson *et al.* 1988; Partanen *et al.* 1987). For example, in a study of RA patients with Felty's Syndrome (an uncommon complication of rheumatoid disease), Thomson *et al.* (1988) found a significant increase in the C4BQ0 allele, relative to RA patients without the syndrome. Felty's syndrome is probably multifactorial (Pinals, 1981), but circulating immune complexes are thought to be involved (Breedveld *et al.* 1985).

Another example is gold salt therapy-induced pneumonitis, a rare complication in RA patients. Partanen *et al.* (1987) found a high frequency of the C4BQ0 allele in both this sub-group, and other RA patients. This allele occurred most frequently with HLA-B35 and/or HLA-B40.

The previous two studies included few (<20) patients in the sub-groups analyzed, and C4 null alleles were identified by visual assessment of C4 bands, with corroboration in only some cases by family studies. Only Thomson's group used CPseR, and then, only

to repeat samples for which electrophoretic band patterns were difficult to interpret. Furthermore, other investigators have not reported any increase in C4B null in RA (Laurent and Welsh, 1983; Kay *et al.* 1983). Thus, the association of RA with particular sub-types of DR4 remains stronger than the association with any other antigen(s), and particular complement alleles may, again, serve as markers for high risk haplotypes.

Chapter 2

AIMS & OBJECTIVES

The aim of this thesis was to assess the usefulness of densitometry for detecting null C4 alleles, in populations of unrelated individuals.

The rationale for this study was as follows. There are numerous reports of an association between null C4 alleles and certain autoimmune diseases. Clearly, the strength of these associations is dependent on the accurate identification of null C4 alleles. This requires a reliable method which will do this. However, assigning null alleles with certainty to individuals has always posed a problem in disease association studies. One difficulty is the complex banding pattern which results from immunofixation electrophoresis of C4. A second, is that in heterozygous individuals, a null allele may be masked by the allele on the other chromosome. For example, a person whose serum is phenotypically C4A3B1, could be genotypically A3B1/A3BQ0, A3B1/AQ0B1, A3B1/A3B1 or A3BQ0/AQ0B1. The genotype might also be AQ0BQ0/A3B1, however, the AQ0BQ0 haplotype is exceedingly rare.

Family studies have been the best way of assigning null C4 alleles, but often, family data are either not informative, not available, or limited in availability. Moreover, even with family studies, null alleles can only be assigned with certainty to individuals whose C4 types or whose relatives' C4 types are homozygous, or otherwise informative for C4A or C4B. Recently developed molecular genetics techniques are proving to be the most accurate means of identifying deleted C4 alleles, but these techniques still miss those 'nulls' which are not due to gene deletion. These methods

are also time consuming and somewhat labor-intensive. Consequently, relatively small numbers of samples are analyzed at one time.

Densitometry is an attractive alternative to these methods because it is relatively quick, making it useful for analyzing large numbers of samples, the ideal in population studies. Although densitometry has been used for the purpose of identifying null C4 alleles, no systematic studies have been done to assess the accuracy of this method.

The objectives of this thesis were:

(1) To develop a simple and reliable densitometric method to determine relative amounts of C4A versus C4B in individual serum samples

(2) To use this method to address the following questions:

- Is there a relationship between relative levels of C4A versus C4B and number of genes using samples of known genotype, derived from family segregation studies?

- Can relative levels of C4A versus C4B be reliably used to predict an individual's genotype when no family data are available?

- Do other immune complex disease populations, such as rheumatoid arthritics, show a pattern of C4A:C4B similar to SLE populations or control populations?

The study was planned in two parts. Initially, samples from individuals known to have null C4 alleles by family studies, or likely to be carrying null alleles because of MHC associations, would be selected. Samples from individuals without null alleles would also be selected. The C4A:C4B densitometric ratio for these genotyped, informative individuals would be determined after immunofixation electrophoresis of serum treated with carboxypeptidase-B and neuraminidase. The data would then be analyzed by discriminant analysis to ascertain whether persons without null genes had mean C4A:C4B ratios approaching 1; those having one null C4A gene had mean ratios approaching 0.5; and those having one null C4B gene had mean ratios approaching 2.

The predicted gene number based on densitometry would then be compared to the gene number according to family data, in order to assess how well the two agreed.

In the second part of the study, a large number of samples from unrelated individuals belonging to different populations would be analyzed by the densitometry-discriminant analysis method. The aim was to apply the method to populations expected to be quite different from each other, with respect to the distribution of null alleles. Thus, the populations included SLE samples for which an association with C4A-null is reported, and RA samples for which an association with C4B-null has occassionally been reported.

The reasoning at the outset of this study was that using densitometry to determine an individual's genotype would likely be too error-prone, but that it could still usefully be applied to populations in order to compare genotype distributions. The error of the method should be the same for each population studied, thus relative differences should still be apparent, and the analysis would show whether SLE and RA populations, for example, differed from control populations in their distribution of null C4A and C4B alleles.

Chapter 3

MATERIALS & METHODS

This study was planned in two parts. The first part of the study dealt with a panel of genotyped individuals, about whom family pedigrees, haplotypes, HLA-data for -A,-B,-C and -DR antigens and genotypes were known. The second part of the study was based on individuals about whom less was known. These individuals thus represented the type of subjects to whom densitometry would normally be applied, in that they were basically "unknowns". These individuals belonged to three different geographical populations, and for the majority, the only known information was C4 phenotype and the densitometry results from this study. HLA-B and HLA-DR phenotype data were known for some members of each population.

Throughout the text, the genotyped panel will be referred to as genotyped subjects. All other individuals will be referred to as phenotyped population members.

3.1. Subjects

Samples used in this study, had been previously collected as EDTA-plasma or serum, and were stored frozen at -70°C. Prior to use, samples were thawed quickly in a 37°C water-bath, then placed on ice. After use they were re-frozen at -70°C.

The first part of this study was based on the genotyped panel. Serum samples for these subjects were selected from families typed by the Clinical Immunology Diagnostic Laboratory (Health Sciences Centre, Memorial University of Newfoundland), for HLA-A, B, C, and DR. C4-typing was done in the Immunology Research

Laboratories (Dr. V. Skanes). Genotyping was performed by Dr. B. Larsen in the course of various family studies of: healthy lab personnel, transplant recipients and paternity cases, diabetics, individuals with ankylosing spondylitis, and those with Graves' disease.

The second part of the study was based on individuals from three different phenotyped populations: Newfoundland, Canada; St. Louis, USA; and Budapest, Hungary.

St. Louis controls were healthy unselected Caucasian plasmapheresis donors. St. Louis SLE patient samples came from the Washington University Rheumatology Clinics. All SLE patients and controls were Caucasians and all SLE patients satisfied classification criteria of the American Rheumatism Association for SLE.

The Hungarian control serum samples were obtained from medical lab personnel and their spouses. Hungarian SLE samples included patients with both renal and non-renal involvement in the disease. All Hungarian samples were provided by Dr. V. Stensky of the National Institute of Haematology and Blood Transfusion, Budapest, Hungary.

Newfoundland population control samples came from medical lab and hospital personnel; from paternity testing cases; and relatives of potential kidney or bone marrow transplant donors, (samples were limited to one relative per family). Newfoundland RA samples were obtained through the courtesy of Dr. B. Larsen. All patients satisfied the American Rheumatism Association (1984) criteria for classical or definite rheumatoid arthritis.

3.2. C4 genotyping

C4 genotyping of panel subjects was based on C4-phenotype and HLA-haplotype data. Genotyping was performed by the Immunology Research Laboratory (Health Sciences Centre, Memorial University of Newfoundland). About 20 of the 108 genotyped subjects were genotyped with the aid of HLA-association data, because family data were not completely informative. For example, C4AQ0B1 occurs in tight linkage

disequilibrium with HLA-B8, DR3, in all Caucasian populations tested (Awdeh *et al.* 1983). Thus, in cases where the C4A allele could not by deduced but occurred on the HLA-B8, DR3 haplotype, it was classified as C4AQ0. In this laboratory, a haplotype assigned in this way is scored with (*) to indicate the less likely possibility of another allele at the indicated locus. For example, AQ0*B1 means that AQ0 was assigned because it is in strong linkage disequilibrium with the HLA antigens on that haplotype, but a possible, although less likely alternative is A3.

3.3. Complement allotyping

C4 allotyping of EDTA-plasma or serum samples from all three populations, was performed at the Immunology Research Laboratory, (Health Sciences Centre, Memorial University of Newfoundland).

All samples were subjected to IFE again, prior to densitometric analysis. Thus, it was possible to check the phenotypes obtained from these runs against the original recorded genotype or phenotype for each sample; (except for the Hungarian control samples, which were typed for the first time for use in this thesis). Any discrepancies were resolved by repeating the sample run. Where more than one aliquot of the sample was available, each aliquot was run and checked. The family pedigree was re-evaluated for genotyped controls. If a discrepancy remained, the sample was not analyzed.

3.4. Immunofixation electrophoresis

The C4-typing method used in this study, was a combination and a modification of two techniques: the standard method of Awdeh and Alper (1980), and a recent modification of this method by Sim and Cross (1986). In addition, the agarose gels used for C4-typing were cast on GelBond film (FMC Marine Colloids Division, Rockland, Maine USA 04841), rather than glass plates as used conventionally.

Accurate readings using densitometry depend upon well separated and well defined bands. As samples were run on a GelBond film support rather than on conventional glass plates, some modifications had to be made. Because the dimensions of the GelBond film were smaller than those of the glass plates normally used, adjustments had to be made in: the volume of agarose used, the pattern used for sample application to a gel, and the voltage used during the run. Therefore, a series of preliminary runs, with variations described in the following sections, were done to establish the best conditions for electrophoresis using GelBond film.

3.4.1. Casting the agarose gel

In a modification of two methods, "open casting" and "moulding" (FMC Marine Colloids Division product information booklet, BioProducts Department, Rockland, Maine 04841, 1980), GelBond film was placed, hydrophilic side up, on a glass plate of approximately the same dimensions as the film (21 cm x 16 cm). A Plexiglass frame was clipped over this combination, and the assembly was levelled on an LKB plate-leveller [Pharmacia (Canada) Inc., 2044 St. Regis Boulevard, Quebec, H9P 1H6]. Two frames were tested. The first measured (20 x 20 x 0.1) cm, yielding a frame with final inner dimensions of (18 x 18 x 0.1) cm. The second measured (21 x 16 x 0.4) cm, yielding a frame with final inner dimensions of (19 x 14 x 0.4) cm. Of the two frames tested, the one with the smaller width, but greater thickness was chosen, (21 x 16 x 0.4) cm. The larger frame proved too fragile, and cracked easily when it was clipped to the glass support plate. Also, the extra width was unnecessary for the duration of the electrophoresis run. To prepare the gel, 0.5% agarose (Type I: Low EEO, No. A-6877, Sigma Chemical Company, P.O. Box 14508 St. Louis, Mo. USA 63178) in gel buffer (7.9 mM sodium-barbitone, 1.4 mM barbitone, 93.6 mM glycine, 46.6 mM Tris, 50.0 mM disodium EDTA), was poured onto the GelBond film. When both the smaller and the larger frames were being tested, 27 ml or 32 ml, respectively, of 0.5% agarose was used. In each case

the final gel thickness was 0.1 cm. The agarose solution was allowed to set at room temperature for several minutes. If the gel was not used immediately, it was stored at 4°C in a humidified container for a maximum of two days. Three types of agarose were tested to determine which one produced the best run with respect to band separation and definition: Type II Medium electro-endo-osmosis (EEO) (Sigma No. A-6877); Type I: Low EEO (Sigma No. A-6013) and Type L-Behring (Hoechst Canada Inc. Behring Diagnostics, 4045 Côte Vertu, Montreal, Quebec H4R 1R6). Using Type II Medium EEO agarose, the bands produced were more diffuse than with the other two types tried, and the samples had to electrophorese longer to obtain separation comparable with those two. Type L-Behring agarose and Type I Low EEO agarose produced equally sharp, well separated bands for the same amount of running time. Type I Low EEO agarose was selected for use.

GelBond is a flexible support medium for agarose gel, consisting of agarose-coated polyester. Dried Gelbond could be cut into strips and conveniently read using the densitometer. GelBond had other advantages. Because it was light-weight and pliable, it could be easily inserted into the sample carrier, and later taped above the densitometer printout, as a permanent record. Also, by doing this, a visual impression of staining intensity of the C4A and C4B bands could be compared with their respective curves on the printout. This was a good method to ensure that the sample ratio had not been read or calculated "backwards" (i.e. C4B:C4A).

3.4.2. Treating serum samples

Initially, 20 μl of serum was diluted in C4 Sample Buffer (20mM EDTA, PBS, 2.0 mM PMSF) at 1:3, 1:4, 1:5 and 1:8. A serum dilution of 1:4 in C4 Sample Buffer produced the best results during preliminary runs. The samples tested at the various dilutions had been collected recently, frozen at -70°C, and thawed for only the first or second time. However, during subsequent runs of test samples, which were various ages and had unknown freeze-thaw histories, 1:4 was found to be too dilute for many

samples. The reason for this is not known. Most samples in the genotyped panel and in all populations were diluted at 1:2. The optimum dilution for samples which did not run well at 1:2 was gauged from the dilution results for the preliminary results.

C4 Sample Buffer was used to dilute serum samples for several reasons. The EDTA it contained chelated calcium, preventing activation of the classical complement pathway and PMSF inhibited proteases which are minor contaminants of neuraminidase and carboxypeptidase-B. These contaminants may produce extra bands in some samples if they are not inhibited.

Sim and Cross recently (1986) introduced the carboxypeptidase-B (CPseB) modification to the standard immunofixation electrophoresis technique of Awdeh and Alper (1980). CPseB treatment of serum prior to IFE has advantages for densitometry and improves the reliability of densitometric results. C4A:C4B densitometric ratios are unaffected by CPseB treatment (Zhang *et al.* 1988). Allotype assignments are easier because of the removal of overlapping triplets, duplicated alleles such as C4A3A2 are more obvious, and it is possible to distinguish between alleles with small differences in electrophoretic mobility, such as C4A1 and C4B3. All these effects improve the correct assignment of C4 alleles, which is necessary for accurate calculations of C4A:C4B ratios by densitometry. The method used in this study differed from that of Sim and Cross in the following respects:

(1) Duration of CPseB incubation

(2) Concentration of CPseB used

(3) Use of C4 Sample Buffer to dilute serum

(4) Type of neuraminidase used

During preliminary runs in which the dilution of serum in C4 Sample Buffer was varied, diluted samples were incubated for 30 minutes with 0.13 units of CPseB (Sigma Type I-DFP EC.3.4.17.2, No. C-7261) per μl of serum, as per the method of Sim and Cross (1986). Incubation times of 10 minutes and 5 minutes were then tried, and

it was found that the banding patterns were as good as those obtained after a 30 minute incubation. Thus, an incubation time of 10 minutes was chosen for convenience.

Three units of CPseB were tested at dilutions of 1:1-1:10 inclusive, in serum diluted 1:2 in C4 Sample Buffer. All dilutions up to 1:6 produced equally good banding patterns. Those from 1:7-1:10 showed a faint extra band in the C4B region. From these results, it was decided to use CPseB diluted 1:4 in PBS.

Serum samples (20 μl) were thus treated with 4 μl of CPseB which had been diluted 1:4 in PBS (0.15 units CPseB/μl enzyme solution). The final concentration was 0.03 units CPseB/μl serum. This was a four-fold reduction in the amount used by Sim and Cross.

The digest was then treated with 0.16 units of neuraminidase Type V (Sigma No. N-2876), and left to incubate overnight at room temperature. Sim and Cross used Type VIII neuraminidase (Sigma No. N-5631), which has fewer enzyme impurities than Type V. Both enzymes were tested at the concentration specified above, and no difference was found in the pattern produced by samples treated with Type V versus Type VIII. Type V, being more ecomomical, was chosen for use.

3.4.3. Electrophoresis

To load treated serum samples onto the agarose gel for electrophoresis, sample slots were made by manually inserting strips of filter paper (Whatman chromatography paper 3 mm, Whatman International Ltd., Maidstone England) into the gel, 3 cm from the cathodal edge.

Two different slot widths and two different sample volumes were evaluated. 5 μl of sample was applied to 5 mm and 7 mm slots, and 6.5 μl of sample was applied to 5 mm and 7 mm slots. Initially, 20 samples were applied to a gel. Good electrophoretic patterns were produced from 7 mm slots containing 6.5 μl of treated serum. For this slot size and the width of the GelBond used, it was found that 14 was an ideal number of samples to load on each gel.

The heat generated during electrophoresis proved to be a problem in the early runs. The agarose-coated polyester of GelBond is a poor heat conductor compared to the glass plates normally used for electrophoresis. Consequently, the heat in the gel could not be transferred to the cold water of the tank bed. If too much heat was generated, the gel cracked, terminating the run. (Preliminary runs were carried out at a voltage close to that used for glass plates). Preliminary gels were compared using constant current (50mA); constant voltage (550v or 525v or 500v); or constant power (35W), (Buchler 3-1500 power supply, Buchler Instruments, Fort Lee, New Jersey USA). Comparing results for constant voltage, constant current and constant power, it was found that gels were best run at constant voltage. Of the three voltages tested, the heating problem was eliminated at 500v. Samples were therefore subjected to electrophoresis at a constant voltage of 500v for about three hours. The electrode buffer contained 31.8 mM sodium-barbitone, 5.6 mM barbitone, 374.3 mM glycine, and 186.6 mM Tris. A hemoglobin marker was used to estimate the distance samples had migrated during the run. The electrophoresis apparatus (Dept. of Technical Services, Memorial University of Newfoundland, Canada) was similar to the LKB Bromma 2117 electrophoresis unit [Pharmacia (Canada) Inc.] and was cooled by tap water. During summer months when the temperature of the tap water cooling the electrophoresis apparatus increased, the voltage was reduced to 450v.

3.4.4. Immunofixation

A sporadic feature of early runs was bands that had areas of weak staining in the centre of the band. The possibility that this was caused by using different lots of antibody was checked. Two different lots of goat anti-(human) C4 serum were used to detect C4 bands: Atlantic Antibodies anti-human C4 complement (B1E, Lot 058), and Atlantic Antibodies anti-human C4 complement (B1E Lot 033), (Atlantic Antibodies P.O. Box 60, 10 Nonesuch Road, Scarborough, Maine USA 04074). However, both lots gave similar results.

The possibility that the weak staining areas were caused by antigen or antibody excess was checked by varying the amount of anti-C4 spread on the gel. Initially, 250 µl in 1.5 ml PBS was used. When the amount of antibody was reduced to 150 µl in 1.5 ml PBS, the weak staining areas were no longer present.

After electrophoresis, goat anti-(human) C4 serum (Atlantic Antibodies, Lot 033 or 058) at a concentration of 0.5U/cm^2 was applied to the gel. The gel was then incubated in a humidified container at 37°C for one hour. Following incubation, the gel was pressed for 5 minutes with a piece of Whatman #3 filter paper soaked in 0.9% sodium chloride (NaCl), under a 5 kg lead weight. The gel was soaked overnight in a 0.9% NaCl solution to remove unprecipitated proteins. This was followed by a rinse in distilled water for 15 minutes. The gel was then dried briefly in a 90°C oven (Thelco Model 15, GCA Precision Scientific, Chicago, Illinois USA 60647).

The dried gel was placed in 0.5% Coomassie Blue stain (Sigma) for fifteen minutes. It was then rinsed in water, placed in a destaining solution (150 ml glacial acetic acid, 450 ml absolute ethanol, 450 ml distilled water) for five minutes, rinsed in water once more, then left to dry at room temperature.

3.5. Densitometric analysis of C4A and C4B protein bands

Having established the necessary modifications for IFE using GelBond, the next step was to check the operation of the densitometer which would be used in this study, and standardize the procedure for its use. The densitometer is a scanning device which can be used for direct or indirect quantitation of substances separated electrophoretically e.g. serum proteins. The method used in the current study was indirect, in that the separated C4 bands first had to be stained with Coomassie blue. Densitometry was done using an ACD-18 Automatic Computing Densitometer (Gelman Instrument Company, Ann Arbor Mich.). The absorbance of the bands was measured by the densitometer, with settings optimal for the Coomassie Blue counter-stain:

wavelength 595 nm, slit width 0.2 x 3 mm, autogain 80%, and scan length, 30 mm. Total protein was specified as 1. Background was set to zero before bands were scanned. The ACD-18 densitometer had an optical density (O.D.) range of 0.25-6.0 O.D., which was more than adequate for the range of C4 protein concentrations encountered in this study.

To analyze samples densitometrically, the dried GelBond film was cut into uniform strips corresponding to each sample track on the gel. The ACD-18 operator's manual provided no information on scanning pliable GelBond strips, therefore to determine the most convenient way to load the samples in the sample carrier, a variety of "mountings" were tried. The strips were alternately taped to a glass slide, taped to an acetate square, inserted into a plastic evelope, or taped flat into the carrier. Of the various ways in which GelBond strips were inserted into the carrier to be read, taping just the strip into the carrier, flat and parallel to the carrier top, proved best. The centre of each strip was marked (to aid in making consistent readings), and the strip was inserted into the sample carrier, so that the order of scanning was C4A band first, C4B band, second. The densitometer plotted transmitted light from the stained GelBond strip versus distance along the strip in the form of electrophoretic curves (Figure 3.1). The area under the curves was determined automatically by scaled integration. This area was equivalent to the density of each protein band, and was expressed as a fraction of the total protein, which was specified as 1. The ratio of C4A:C4B was calculated for each sample, as shown in Figure 3.1.

3.5.1. Precision of densitometer

The precision of the densitometric readings was checked as follows. Samples were read in each of ten different tracks of the sample carrier; one sample was read at random intervals over several months; and sample scans using each of the five available slit widths were compared. To check the consistency of the operator's method, one sample was inserted, scanned, removed, then re-inserted. This was done ten times. To determine if it was possible to read several strips of the GelBond at one time, ten

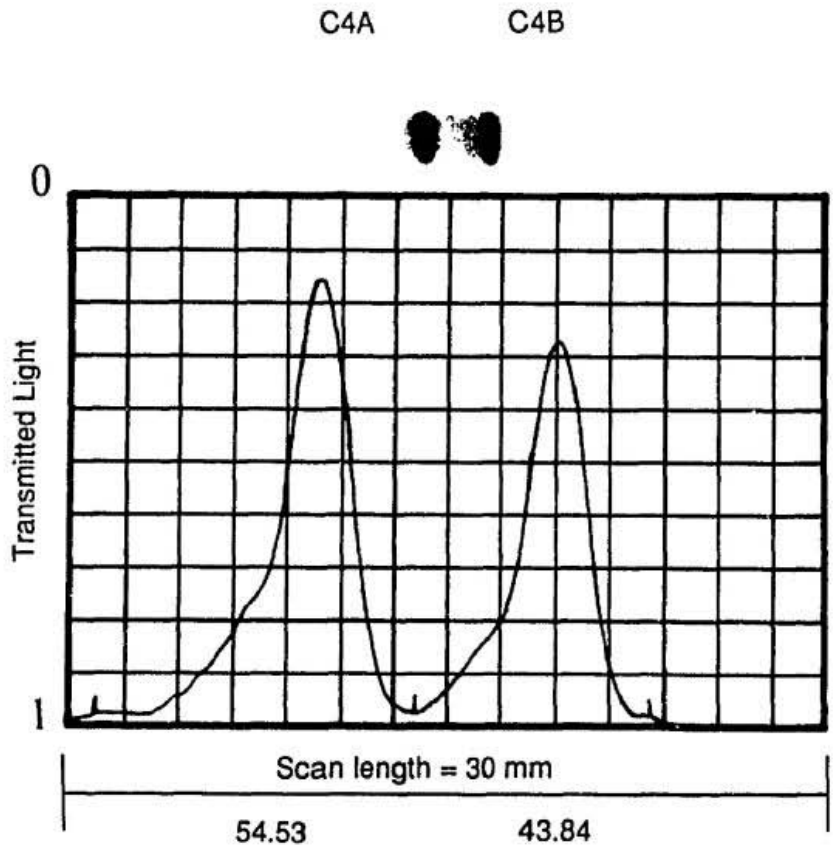


Figure 3.1 **Densitometric results for a sample with 2 C4A and 2 C4B genes (C4A3B1/C4A3B1) by family studies.** The Gelbond strip was scanned over a length of 30 mm and the densitometric output was as illustrated above.

$$\text{The C4A:C4B ratio} = \frac{\text{Area under curve A}}{\text{Area under curve B}} = \frac{54.53}{43.84}$$

$$= 1.24$$

samples were inserted and read consecutively. The samples used in each of these tests were randomly chosen from the genotyped controls.

Having worked out the technical aspects of the densitometric method, subjects belonging to the genotyped panel, and members of the three phenotyped populations, were analyzed densitometrically. After obtaining C4A:C4B ratios by densitometry, but before the total data were analyzed, some samples were excluded. Specifically, for all subjects, those whose C4A and C4B bands were too close together electrophoretically to give separate peaks on the densitometer printout (eg. C4A3B4), were excluded. Homozygous null C4A or C4B samples were excluded. Samples that had degraded, as evidenced by diffuse or streaked protein bands, were also excluded. Among genotyped subjects, a few were found to be incorrectly typed in the original family file, and similarly, some members of the phenotyped population showed phenotypes which were inconsistent with the original C4-typing for them. These were also excluded from the data analysis.

3.5.2. Within and between-run reproducibility

To determine within-run precision, two different samples, I and II, with expected ratios of 1 and 2, respectively, were applied to all the lanes (14) of two gels. To obtain between-run imprecision, two other samples, III and IV, with expected ratios of 0.5 and 1, respectivley, were run in the same sample slot positions on twelve random gels, over a period of three months. To eliminate any possible effects of repeated thawing and freezing, aliquots of each sample used for between-run precision were stored beforehand. The C4A:C4B ratios and the co-efficient of variation (CV) for each reproducibility test were calculated and compared.

3.6. Data analysis

3.6.1. The genotyped panel

The first samples analyzed by densitometry were those from the genotyped panel. The individuals in this panel had been specifically selected for their genotype according to family studies. At the beginning of the study, they were assigned to one of three groups based on their relative gene number. The assignment was as follows. Individuals known to have one null C4A allele were assigned to group 1 (3-gene individuals, heterozygous null for C4AQ0). Those known to be without null genes (4-gene individuals), or hemizygous null for C4A and C4B (2-gene individuals), were assigned to group 2. Those known to have one null C4B allele (3-gene individuals, heterozygous null for C4B) were assigned to group 3. For convenience, group 2 subjects will sometimes be referred to as non-null, with the possibility of hemizygosity for null alleles understood. The C4A:C4B densitometric ratio was calculated for 108 genotyped subjects, then recorded under the subject's group according to family studies. The mean A:B ratio and standard deviation for each of the three groups was then calculated using the statistical program "Autoexec" (copyright David W. Clarke, Queen's University, Kingston Ontario, Canada, 1986). The range of A:B ratios for each group was also determined. (This range will be referred to as the "real range" to distinguish it from the range assigned by discriminant analysis).

As Coomassie Brilliant Blue has linear staining characteristics (Christiansen *et al.* 1983), the simplest hypothesis is that a densitometric ratio of 0.5 or 2.0 represents heterozygosity for C4A null (group 1) and C4B null (group 3), respectively, while a ratio of 1.0 should indicate an equal number of alleles at each locus (group 2). These values then, represented the predicted mean ratios for each group. A Kruskal-Wallis one-way analysis of variance (ANOVA) was performed (Siegel, 1956), to determine if there was a significant relationship between C4A:C4B ratio and group by family studies. The

means of the three groups were then compared to ascertain whether they approached the predicted value for each group.

Next, the A:B ratio data for the genotyped panel were evaluated by discriminant analysis using the statistical package SYSTAT (Wilkinson, Leland. SYSTAT: The System for Statistics. Evanston, Illinois: SYSTAT Inc., 1978).

The purpose of the discriminant analysis was to examine the value of densitometry as a predictor of group 1, 2 or 3 membership i.e. genotype in terms of number of null alleles. Discriminant analysis is a type of analysis which predicts group membership with the particular purpose of maximally discriminating among groups. It is especially useful when there is a possibility of membership in two or more groups, as was the case in this study. The data entered into the program for each genotyped subject, were group (1, 2 or 3) by family studies, and A:B ratio by densitometry. The analysis converted each A:B ratio into a score ("Factor" in Appendix A, p.128), and set up a distribution of scores for each group. A "cutting score" was determined based on the mean of each group. These were values for the A:B ratios where the smallest number of incorrect decisions about group membership would be made. Here, there were two cutting scores, one distinguishing group 1 from group 2, and another distinguishing group 2 from group 3. For example, if a subject's score was greater than the cutting score for group 1, but less than the cutting score for group 3, he/she would be predicted to be a member of group 2. (Cutting scores are not shown in the program output). Discriminant analysis maximizes the accuracy of prediction of group membership by setting up a distribution of scores such that there is minimum overlap between groups.

The range for each group was deduced by looking at the ratios of subjects assigned to groups 1, 2 and 3 by discriminant analysis of densitometric ratios, and finding the highest and lowest ratio in each of these groups. The program calculated discrete probabilities (Appendix A) for a subject's membership in each group by minimizing the overlap between groups. In simplified terms, calculation of probabilities by the program was based on comparing the 'distance' each ratio was from the mean of each group, so

that a probability for membership in group 1, 2 and 3 was derived for each subject (Appendix A). (For a detailed description of discriminant analysis see Thorndike, pp.203-220, 1978).

The results of the discriminant analysis were then used to generate a "truth-table". This table summarized the results of group assignment based on family studies ("group", Appendix A) as compared to group predicted from A:B ratio by densitometry ("predict", Appendix A).

In order to interpret the truth-table results, the data in this table were transformed into 2 x 2 tables, providing an easy format for computing the proportion of subjects with any particular result. Using the data in the 2 x 2 tables, the method was evaluated in terms of accuracy, true-positive, true-negative, false-positive, and false-negative values (Fletcher, Fletcher and Wagner, 1982).

Following the discriminant analysis, genotyped subjects whose group as determined by densitometry differed from that as determined by family studies, those with nearly equal probabilities of being in either of two groups, and those with very high or very low A:B ratios, were examined more critically to determine whether or not they shared a common feature (e.g. C4-type, HLA-type). As well, A:B ratios for 94 of 108 genotyped panel subjects were evaluated in order to determine whether particular C4 genotypes were associated with high or low ratios. Only genotypes for which there were three or more samples were considered.

3.6.2. Phenotyped population members

The discriminant analysis results from part I were applied to the members of the phenotyped populations, in order to predict group membership for these individuals. The SYSTAT MGLH module was used for this part of the analysis because it had a "weight" command which can be used to assign probabilities for group membership to individuals by comparing their A:B ratios with the range of ratios for the genotyped

panel. Each phenotyped population member was assigned a probability for membership in each of the three groups by the program, and the highest probability determined to which group they were assigned. From these results, the number of subjects in each group was counted to determine observed (O) frequencies. The expected (E) frequency for samples in each group was calculated according to Hardy-Weinberg expectations (Suzuki *et al.* 1986) using the observed frequency of homozygous null C4A and C4B in each population. Observed versus expected frequencies for each phenotyped population were compared using the Chi-square test.

As for the genotyped panel, phenotyped population members with nearly equal probabilities of membership in either of two groups, and those with very high or very low A:B ratios, were selected out for further study.

3.7. Single radial immunodiffusion (SRID)

In order to compare total serum C4 with gene number by family studies, total serum C4 was determined by SRID for 78 randomly selected genotyped subjects (Mancini, Carbonara, and Heremans, 1965), using Behringwerke Nor-Partigen-C4 plates (Hoechst, Canada Inc. Behring Diagnostics, Montreal Quebec). Subjects were divided according to family studies into the following categories: 2 genes (hemizygous null), 3 genes (heterozygous null C4A or C4B), and 4 genes (non-null). Values for heterozygous null A and B were plotted separately to determine if there was any difference between the two. Total serum C4 concentration for these subjects was plotted against gene number according to family data. The statistical package "Epistat" (copyright Tracy L. Gustafson, 1984) was used to calculate the mean total C4 concentration for each genotype group and perform a one-way ANOVA. Samples with the highest and lowest total serum C4 values in each group were evaluated in order to determine whether particular A:B ratios correlated with very high or very low total C4. The hemizygous null group was not considered, as there were only three samples in this group.

Chapter 4

RESULTS

4.1. Exclusion of samples analyzed densitometrically

Of the 613 samples analyzed densitometrically (Table 4.1), 61 (10%) were excluded from the final data analysis for the reasons outlined below (Table 4.2).

Among genotyped controls, a few samples were excluded where C4-typing results were inconsistent with original family data, and this was not resolved by repeating electrophoresis or checking the family pedigree. It should be noted that this did not include samples typed as C4A3,A2 using CPseB treatment, which may originally have typed as C4A3, but only samples which were a significantly different allotype compared to the original typing. Some of the samples were received as aliquots from the original tubes, therefore mis-labelling may have occurred in some cases. With the exception of the Hungarian control population, only a small number of phenotyped samples from each population were excluded from analysis of the densitometric results (Table 4.2). Of these, nearly one-third were excluded because of overlapping peaks on the (densitometric) printout indicating that, on the gel, the C4A and C4B bands were very close. This usually corresponded to C4A3,B4 phenotypes, but included some samples phenotyped as C4B3. Densitometry was done on 38 Hungarian controls, but only 24 were included in the data analysis. Of the 14 samples excluded, 2 were C4B4, and 5 were homozygous null C4A or C4B. The remainder could not be reliably identified by name, because labels on the serum tubes were partially or completely rubbed off. The possibility that the same sample might be analyzed twice would not be excluded, as the controls were

Table 4.1

Number of samples analyzed by densitometry in the genotyped panel and in the phenotyped populations

	Controls	Patients
Genotyped panel	119	NA*
Phenotyped populations		
Newfoundland	120	92
Hungarian	38	42
St. Louis	128	74

*NA - not applicable

Table 4.2

Samples analyzed but excluded from discriminant analysis of C4A:C4B ratios

Population	No. analysed	No. excluded	Reason excluded: Overlapping peaks on printout	Reason excluded: Other[a]	Reason excluded: Homozygous null C4A or C4B
Genotyped panel	119	11 (9%)	3	8	
Newfoundland controls	120	12 (10%)	2	3	7
Newfoundland patients	92	12 (13%)	1	-	11
St. Louis controls	128	6 (5%)	5	1	
St. Louis patients	74	4 (5%)	4	-	
Hungarian controls	38	14 (38%)	2	7	5
Hungarian patients	42	2 (5%)	1	-	1

[a]other - poor or untypable samples

- for genotyped panel this included subjects whose C4-typing results were significantly at odds with original records

- incomplete labels on tubes

supplied in at least duplicate. Samples which showed evidence of degradation (diffuse or streaked protein bands; un-typable samples) were excluded on a subjective basis from all populations. Lastly, homozygous null samples were not included in the data analysis because an A:B ratio cannot be calculated for such samples.

4.2. Precision of densitometer

The precision of the readings obtained from the densitometer used for this study is shown in Tables 4.3-4.8. Table 4.3 shows that the track from which samples were read made little difference. However, when an attempt was made to read ten samples in consecutive tracks, it was discovered that the machine was reading the first sample, then re-scanning the edge of this sample closest to the next sample, which was then recorded as the reading for sample no. 2. This could be prevented by increasing the spacing and scanning only a few samples at a time, but it was actually quicker to scan samples one at a time, rather than align several, tape them into the carrier, etc. All samples were read using a scan length of 30 mm, so that the total area scanned for each sample was consistent.

Readings for the same sample scanned using the five different slit widths available, varied somewhat. Several samples were compared this way, and the results of a representative sample are shown in Table 4.4. Using five different slit widths, the A:B ratio for this sample varied from 0.27-0.45. The slit width which produced a ratio closest to the average of the five was chosen for reading all samples. This was 0.2 x 3 mm. The operator's method also proved to be consistent (Table 4.5). Finally, the ratios obtained for a sample read at random over several months varied little (Table 4.6). This was done as a check on the general operation of the machine over time.

TABLE 4.3

Precision of densitometer for sample #5 read in ten different tracks with slit width 0.2 x 3 mm

Track (n)	C4A : C4B Ratio (x)
1	1.12
2	1.11
3	1.08
4	1.10
5	1.09
6	1.09
7	1.08
8	1.08
9	1.09
10	1.10
Mean (x)	1.09
SD*	0.01

* calculated using $SD = \sqrt{\frac{(x - x)^2}{n - 1}}$

TABLE 4.4

Precision of densitometer using different slit widths and sample #6

Slit Width (mm)	C4A : C4B Ratio
0.1 x 2	0.45
0.1 x 5	0.45
0.2 x 3	0.40
0.2 x 10	0.27
0.5 x 3	0.43
Mean	0.40

TABLE 4.5

Consistency in operator's technique for ten densitometric readings[a] in the same track with slit width 0.2 x 3 mm

	C4A : C4B ratio
	1.10
	1.10
	1.10
	1.08
	1.09
	1.07
	1.04
	1.05
	1.10
	1.09
Mean	1.08
SD	0.03

[a] sample #7

TABLE 4.6

Precision among densitometric readings for sample #8 scanned at random over several months

	C4A : C4B Ratio
	1.00
	1.14
	1.29
	1.22
	1.13
Mean	1.16
SD	0.11

TABLE 4.7

Within-run precision of densitometric measurements for two samples

	C4A:C4B ratio	
	I	II
	1.13	1.87
	1.15	1.85
	1.08	1.80
	1.07	1.71
	1.09	1.68
	1.15	1.65
	1.12	1.65
	1.23	1.59
	1.21	1.58
	1.17	1.53
	1.12	1.53
	1.10	1.58
	1.07	1.54
	1.11	1.64
Mean A:B ratio	1.13	1.66
SD	0.05	0.11
CV (%)[a]	4.2	6.6

[a] co-efficient of variation - the standard deviation expressed as a percentage of the mean

TABLE 4.8

Between-run precision of densitometric measurements for two samples

	C4A:C4B ratio	
	III	IV
	0.47	1.10
	0.41	1.00
	0.44	1.37
	0.50	1.24
	0.37	1.11
	0.46	1.11
	0.48	1.07
	0.47	1.27
	0.58	1.15
	0.39	1.02
	0.57	1.23
	0.47	1.32
Mean A:B ratio	0.47	1.17
SD	0.06	0.12
CV (%)	13.2	10.3

4.3. Within and between-run reproducibility

4.3.1. Within-run precision

Within runs, the co-efficients of variation for the C4A:C4B ratios were less than 10% and 5% (Table 4.7). Most of the C4A:C4B ratios were within mean ± 1SD, and all were within mean ± 2SD.

4.3.2. Between-run precision

Ratios between runs varied more than those within-runs, with co-efficients of variation between 10% and 14% (Table 4.8). Again, most of the C4A:C4B ratios were within mean ± 1 SD, and all were within mean ± 2SD.

The co-efficient of variation (CV) values obtained for within and between-run variability were comparable to those reported for densitometric analysis of other serum proteins (Dennis, Biegler and Papas, 1986). Also, the variation within and between-runs was low enough for the samples tested, that none were shifted into another group e.g. from group 2 to group 1 or 3 for sample 1 (Table 4.7).

4.4. C4A:C4B densitometric ratios for genotyped subjects

4.4.1. Mean C4A:C4B ratios

The means of groups 1, 2 and 3 by family studies, to which the 108 genotyped controls belonged, were compared. Mean ratios did approach the predicted values for each group (p. 43, Table 4.9). For subjects heterozygous null for C4A (group 1), the mean C4A:C4B ratio approached 0.5; for those heterozygous null for C4B (group 3), the mean ratio approached 2; and for those who were hemizygous null or had no null alleles (group 2), the mean ratio approached 1. The standard deviation for each of the three groups was high. Nevertheless, a one-way analysis of variance showed that there was a significant relationship between group by family studies and C4A:C4B ratio ($p < 0.001$).

TABLE 4.9

C4A:C4B densitometric ratios of HLA and complement genotyped subjects

	Group by family studies		
	1 (Heterozygous null A)	2 (Non-null or Hemizygous null)	3 (Heterozygous null B)
n	47	40	21
Real range	0.21-1.14	0.49 - 2.30	0.44 - 3.33[a]
Mean	0.54	0.98	2.06
SD	0.20	0.34	0.73

[a] If the lowest ratio in this group is removed, the next value is 1.12 (see p. 59)

4.4.2. Range of C4A:C4B ratios

Although C4A:C4B ratio was related to group according to family studies, the range of the ratios for each of groups 1, 2 and 3, was broad, resulting in considerable overlap between groups (Table 4.9, Figure 4.1). The greatest overlap occurred between groups 1 and 2. The C4A band for the sample with the lowest ratio (0.44) in group 3 (Table 4.9) looked degraded, and the sample should have been excluded on this basis, but it was missed. As a result, the sample appeared to be mis-classified by densitometry, the overlap between the three groups was increased, the range of ratios for group 3 appeared larger than it actually was (0.44-3.33 compared with 1.12-3.33, if the lowest ratio was excluded), the standard deviation was increased and the mean was lowered.

4.5. Discriminant analysis of C4A:C4B densitometric ratios for genotyped subjects

The output from the discriminant analysis of A:B ratios for the genotyped panel is shown in Appendix A, and explained on pp. 44-45. The range of ratios for each group, as defined by discriminant analysis, was such that there was no overlap between groups (Table 4.10).

The results of the discriminant analysis were used to generate a truth-table (Table 4.11). This table summarized the results of group assignment based on family studies, as compared to group predicted from A:B ratio by densitometry. Thus, it provided a comparison of how well the two agreed, which in turn reflected how useful the densitometric method was for detecting null C4A or C4B alleles.

4.5.1. Interpreting the truth-table.

The data from the truth-table were transformed into 2 x 2 tables (p. 45, Tables 4.12a, 4.13a, 4.14a) and the parameters used to evaluate the method are defined below, in terms of the complement null alleles which were investigated in this project.

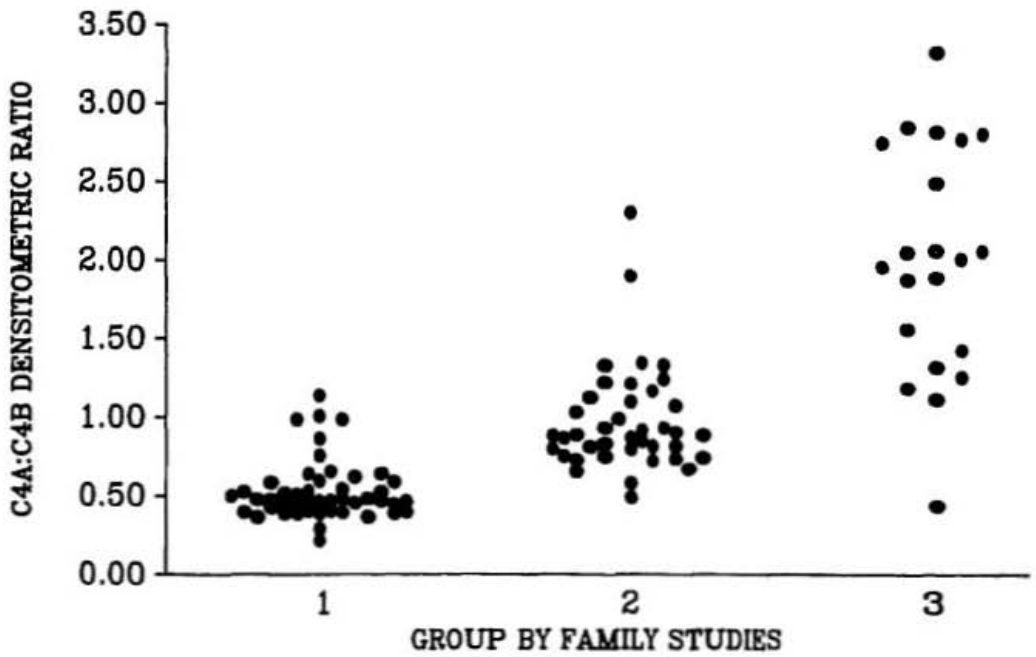


Figure 4.1 **Observed C4A:C4B ratios for genotyped subjects**

Group 1 = heterozygous null for C4A

Group 2 = non-null or hemizygous null

Group 3 = heterozygous null for C4B

TABLE 4.10

A comparison of the real range of C4A:C4B ratios for genotyped controls, with the ranges defined by discriminant analysis

	Group		
	1	2	3
n	47	40	21
Real range by family studies	0.21 - 1.14	0.49 - 2.31	0.44 - 3.33
Range by discriminant analysis	0.21 - 0.76	0.80 - 1.43	1.56 - 3.33

TABLE 4.11

"Truth-Table" generated by discriminant analysis of complement genotyped controls

		Group by Densitometry			
		1	2	3	Total
Group by Family Studies	1	41	6	0	47
	2	10	28	2	40
	3	1	5	15	21
Total		52	39	17	108

4.6. Accuracy of the densitometric method

Accuracy is a property which summarizes the overall value of a test. It is defined as the proportion of all results, both positive and negative, which are correct. In the present study, this was the proportion of samples which were assigned to the same group by family studies and discriminant analysis of densitometric results, i.e. were correctly classified. For example, the accuracy of predicting subjects heterozygous null for C4A was 84% (Table 4.12b). The accuracy for predicting C4B null was about 10% higher than this (Table 4.13b), while for group 2 it was lower (79%), (Table 4.14b). Using the C4A null group (1) as an example, the accuracy was calculated as follows (Table 4.12a). Forty-one of 47 subjects identified as heterozygous null A by family studies, were identified as such by densitometry. Therefore, 41 was the number of correct "positive" results. Fifty of 61 subjects who were not null A, were identified as such by densitometry. Therefore, 50 was the total number of correct "negative" results. Adding these, and dividing by the total number of genotyped panel members, gave the accuracy of the method for null C4A (Table 4.12b). Overall, 78% of subjects were assigned by densitometry to their correct group. This value was arrived at by checking the truth-table diagonal for correct assignments, (41+28+15), then dividing by the total (108), (Table 4.11).

4.6.1. The "gold standard"

Assessment of the method's accuracy rests on its relationship to some way of knowing whether a null allele is truly present or not. This is referred to as the "gold standard" (Fletcher *et al.* 1982). Here, the gold standard was the group assigned by family studies.

TABLE 4.12a

Two by two table for C4A null constructed from Table 4.11

		Group by Densitometry		
		1	2+3	Total
Group by Family Studies	1	41	6	47
	2 + 3	11	50	61
Total		52	56	108

TABLE 4.12b

Evaluation of the densitometric method for assigning C4A null

Parameter	Fraction of samples	%
True +	41/47	87
True -	50/61	82
Accuracy	91/108	84
False +	11/61	18
False -	6/47	13

TABLE 4.13a

Two by two table for C4B null constructed from Table 4.11

		Group by Densitometry		
		3	1+2	Total
Group by Family Studies	3	15	6	21
	1 + 2	2	85	87
Total		17	91	108

TABLE 4.13b

Evaluation of the densitometric method for assigning C4B null

Parameter	Fraction of samples	%
True +	15/21	71
True -	85/87	98
Accuracy	100/108	93
False +	2/87	2
False -	6/21	29

TABLE 4.14a

Two by two table for non-null and hemizygous null genotyped panel members constructed from Table 4.11

		Group by Densitometry		
		2	1+3	Total
Group by Family Studies	2	28	12	40
	1 + 3	11	57	68
Total		39	69	108

TABLE 4.14b

Evaluation of the densitometric method for identifying C4 non-null individuals

Parameter	Fraction of samples	%
True +	28/40	70
True -	57/68	84
Accuracy	85/108	79
False +	11/68	16
False -	12/40	30

4.7. True-positive and true-negative results

True-positive results refer to the proportion of individuals assigned a C4 null allele by family studies, who were determined to have a null allele by densitometry. Again, using group 1 C4A null subjects as an example (Tables 4.12a,b), 47 were heterozygous null A by family studies, and 41 of these were identified as such by densitometry. Therefore, the proportion of true-positive results was 41/47 or 87%.

True-negative results refer to the proportion of subjects without a C4 null allele by family studies, who were identified as lacking a null allele by densitometry.

4.8. False-positive and false-negative results

Subjects without a null allele by family studies, but identified by densitometry as having a null allele, made up the false-positive category. The proportion of subjects with a null allele by family studies but not by densitometry, constituted the false-negative category. Together, these represented mis-classified subjects. Using heterozygous null A subjects as an example (Tables 4.12a,b), the false-positive and false-negative proportions were calculated as follows. The total number of subjects without an A-null allele by family studies was 61. Eleven of these were identified by densitometry as having one A-null allele. Thus, 11/61 or 18% were false-positive for A-null. Forty-seven subjects were heterozygous null for C4A by family studies. Six of these were identified by densitometry as non-A-null. Thus, 6/47 or 13% of panel members were false-negative for A-null. There were considerably more false-positive assignments to the A-null group, than to the B-null group (Tables 4.12b, 4.13b). On the other hand, false-negative results showed the opposite trend, with most false-negative subjects belonging to the 2:1 group.

In summary, the results were:

(1). The overall accuracy of densitometric ratio as a predictor of genotype was, in this analysis, 78% (i.e. 78% of samples were assigned to the same group by both family studies and A:B ratio).

(2). By densitometry, there were 18% false-positives for null A (group 1) and 16% for non-null (group 2), compared with 2% false-positives for null B (group 3), (Tables 4.12b, 4.13b, and 4.14b).

This implies that subjects lacking C4B-null were rarely mis-classified by the method as having C4B-null. In contrast, subjects lacking C4A-null were more likely to be mis-classified by the method as having C4A-null. All of the subjects false-positive for group 1 came from group 2 (Table 4.11). (Recall that the group 3 subject classified densitometrically as belonging to group 1, should have been excluded from the analysis). With respect to null alleles, the most common error in the method was the mis-classification of subjects as heterozygous null for C4A, when according to family studies, they lacked the A-null allele.

(3). There were 13% false-negatives for null A (group 1), compared with 29% false-negatives for null B (group 3), and 30% false-negatives for non-null (group 2), (Tables 4.12b, 4.13b, and 4.14b).

Subjects falsely-negative for B-null and A-null were shifted into group 2 in nearly equal numbers. Most of the subjects falsely-negative for group 2 were shifted into group 1 (Table 4.11).

(4). 24 samples (22%) were "mis-classified" by A:B ratio, and 2/3 of these had lower ratios than predicted from genotype whereas 1/3 had higher ratios (Table 4.11).

(5). When A : B ratio was used to predict null genes by this method, the overall trend was a shift from Group 3 (one null B) to Group 2 (no nulls) and from Group 2 to Group 1 (one null A). More samples were therefore predicted to have null A genes and fewer to have null B genes than their genotypes suggested.

4.9. Mis-classified, extreme and borderline genotyped subjects

Among subjects correctly classified by densitometry, some had very low or high A:B ratios and were thus called "extreme". There were also some subjects with nearly equal probabilities of membership in either of two groups, referred to as "borderline" subjects. Some of these were correctly classified by densitometry, and some were not. C4, HLA-B and DR data were evaluated for mis-classified, extreme and borderline subjects, to determine if particular C4 variants or MHC haplotypes were associated with each category.

4.9.1. Mis-classified genotyped subjects

Table 4.15 shows MHC and C4 haplotype data for mis-classified subjects. Where individuals were members of the same family, this is specified. Samples 1-8 appeared to have higher C4A/lower C4B than expected from their genotypes by family studies. No single MHC haplotype recurred in this group, but HLA-B44 occurred 7/16 times. Samples 9-24 had lower ratios than expected from their genotypes. Of the 32 haplotypes in this group, HLA-B7, C4A3B1, DR2 occurred six times, twice in one family. There were nearly as many HLA-B44-containing haplotypes (5/32). One C4 haplotype, A6B1, occurred in 5 of 32 haplotypes, two being in one family. C4A6B1 occurred almost exclusively with HLA-B17,DR7 and is known to be in linkage disequilibrium with these HLA-antigens (Awdeh *et al.* 1983). In this study, the most frequent extended haplotypes observed in each category are among the most common extended haplotypes in Caucasians (Awdeh *et al.* 1983). Examples are the haplotype HLA-B8, C4AQ0B1, DR3 among

TABLE 4.15

Genotyped subjects for which group by densitometry (GD) did not match group by family studies (GFS)

Sample	C4A:B Ratio	GFS	GD	HLA-haplotype #1				HLA-haplotype #2			
				B	C4A	C4B	DR	B	C4A	C4B	DR
1	0.86	1	2	27	A3*	B1*	1	8	AQ0*	B1	3
2	1.03	1	2	27	A3	B1*	ND	8	AQ0	B1	ND
3[a]	1.14	1	2	44	AQ0	B1	8	44	A3	B1	8
4[a]	1.01	1	2	44	AQ0	B1	8	44	A3	B1	8
5[b]	0.99	1	2	16	A3*	B1*	4	8	AQ0*	B1	3
6[c]	0.99	1	2	44	A3	B1*	1	8	AQ0	B1	4
7[c]	1.91	2	3	44	A3	B1*	1	17	A3	B1*	2
8[c]	2.31	2	3	44	A3	B1*	1	17	A3	B1*	2
9	0.66	2	1	17	A6	B1	?6	27	A3	B1	3
10[d]	0.75+	2	1	40	A4	B2	9	37	A3	B1	2
11	0.58	2	1	17	A6	B1	7	W6	A6	B1	7
12	0.75+	2	1	35	A3	B1	7	7	A3	B1	2
13	0.67	2	1	40	A3	B1	?8	7	A3	B1	2
14	0.74+	2	1	--	A3	B1	ND	7	A3	B1	ND
15	0.49	2	1	44	A3	B1	2	49	A3	B1	5
16[b]	0.74+	2	1	16	A3*	B1*	4	35	A3*	B1*	2
17[b]	0.72+	2	1	16	A3*	B1*	4	7	A3*	B1	2
18[b]	0.73+	2	1	16	A3*	B1*	4	7	A3*	B1*	1
19	1.43	3	2	44	A3	BQ0	4	14	A6	B1	7
20	1.19	3	2	7	A3	BQ0	1	17	A6	B1	7
21	1.12	3	2	62	A3	BQ0	ND	35	A4	B2	1
22	1.32	3	2	44	A3	BQ0	ND	7	A3	B1	ND
23[e]	1.26	3	2	44	A3*	B1	7	18	A3	BQ0	3
24[e]	0.44	3	1	44	A3*	B1	7	18	A3	BQ0	3

a,b,c,d e — indicates members of a family (each letter represents a different family)

* — indicates C4 genotypes deduced from known HLA associations

ND — not determined

\+ — borderline ratios

genotyped subjects in the high A/low B category, and in the opposite category, the haplotype HLA-B7, C4A3B1, DR2 and B44, C4A3BQ0, DR4.

Mis-classifed subjects included several family members, and these family members were usually high A/low B (samples 3 & 4; 6, 7 & 8), or low A/high B (samples 23 & 24). One family was inconsistent (samples 5, 16, 17 & 18).

As mentioned previously, (pp. 32-33), about 18% of all genotyped controls were genotyped with the aid of HLA-association data because family data were not completely informative. Among mis-classified genotyped controls, 11 of the 24 had been genotyped in this way, and are marked with an asterisk in Table 4.15. The asterisk indicates the possibility that a non-null allele could be a null allele, and vice-versa. Therefore, this method of genotyping could have contributed to classification errors. For example, in subject 2 (Table 4.15), if the C4B1* on haplotype #1 were C4BQ0, this would account for the approximately 1:1 ratio. The replacement of C4A* or C4B* with one null allele, would account for every densitometric ratio observed in these subjects, save one. The exception is subject 24, who should have been excluded from the analysis (p. 59).

It should be noted that even if the eleven subjects genotyped by HLA associations were excluded, the overall distribution of subjects mis-classified by densitometry would remain the same. That is, most mis-classified subjects would have shifted from group 2 to 1, while the remainder would have shifted from group 1 to 2, and group 3 to 2.

4.9.2. Extreme genotyped subjects

A subject's A:B ratio was defined as "extreme" when the individual was predicted to be in groups 1 or 3 with a probability of 1, or $p \geq 0.8$ and a ratio proportional in its extremeness to the 2A:1B samples which had $p = 1$. For example, a ratio of C4A:C4B = 0.29 for a group 1 sample was equivalent to a C4A:C4B ratio = 3.5:1, for a group 3 sample. Among genotyped subjects, extreme ratios ranged from a low of 0.21 to a high of 3.33 (Table 4.16). All extreme genotyped subjects were correctly classified by

TABLE 4.16

Genotyped panel subjects with extreme densitometric ratios

Sample	C4 A:B ratio	Group by family studies & densitometry	Haplotype 1			Haplotype 2		
			HLA-B	C4	HLA-DR	HLA-B	C4	HLA-DR
1	2.81	3	44	A3BQ0	NT	35	A4B2	NT
2	2.76	3	44	A3BQ0	4	40	A3B1	1
3	3.33	3	44	A3BQ0	1	40	A4B1	7
4	2.78	3	44	A3BQ0	--	62	A4B2	4
5	2.85	3	44	A3BQ0	NT	35	A3*B1	NT
6	0.29	1	8	AQ0*B1	3	16	A3B1	4
7	0.21	1	8	AQ0B1	3	44	A3B1	7
8[d]	0.36	1	8	AQ0B1	3	40	A4B2	9
9[d]	0.36	1	8	AQ0B1	3	17	A3B1	1

d = members of the same family, see Table 4.15
NT= not tested

densitometry. In extreme samples, one haplotype always carried a null C4 gene, whereas in mis-classified and borderline samples there were also non-null subjects.

Nine of 108 (8%) genotyped controls were extreme (Table 4.16). Of the 9 individuals in the extreme group, samples 1-5 had a densitometric ratio consistent with a very high A or very low B, and all of these had one haplotype which was HLA-B44 and C4A3BQ0.

Four samples (6-9) had extremely low densitometric ratios, and all had one haplotype which was HLA-B8, C4AQ0B1, DR3.

4.9.3. Borderline genotyped subjects

Of the 108 genotyped controls, 14 (or 13%) were classified as borderline, with respect to densitometric results (Table 4.17). A borderline sample was defined as one for which two of its three probability (p) values (by discriminant analysis) differed by less than 0.1. For example, a subject with a p for group 1 = 0.44 and a p for group 2 = 0.50, would be assigned to group 2 by discriminant analysis. However, this subject would be considered borderline for membership in either of the two groups.

Among the 14 borderline subjects, 6 (9-14, Table 4.17) were mis-classified. These six all belonged to the group of ten subjects false-positively assigned to group 1 by densitometry (Table 4.15). This suggests that these borderline subjects had marginally lower A or higher B than expected from genotype.

Eight borderline subjects were assigned to their correct group by densitometry (1-8, Table 4.17). Most (2-8) had marginally lower A or higher B than expected from genotype, while one (sample 1) had marginally higher A or lower B. Therefore, the overall trend for borderline subjects was marginally low A or high B. No C4 or HLA-haplotype was very frequent among borderline subjects.

When subjects were mis-classified as being in group 2 or 3, the classification error was never due to borderline ratios (Table 4.15). In contrast, 6/11 of the subjects

TABLE 4.17

Genotyped panel subjects with borderline densitometric ratios

Sample	C4 A:B ratio	Group by family studies	Group by densitometry	HLA Haplotype 1 B	C4	DR	HLA Haplotype 2 B	C4	DR
1	0.76	1	1	44	AQ0B1	8	44	A3B1	8
2d	0.81	2	2	17	A3B1	1	40	A4B2	9
3	0.81	2	2	44	A4B2	NT	15	A3B1	NT
4	0.80	2	2	8	AQ0B1	3	18	A3BQ0	3
5	0.80	2	2	44	A3B1	9	44	A3B1	8
6	0.82	2	2	-	A3B1	NT	7	A3B1	NT
7	0.81	2	2	62	A3B1	?1	16	A3B1	?9
8	0.82	2	2	38	A3B1	2	37	A3B1	4
9	0.75	2	1	35	A3B1	7	7	A3B1	2
10d	0.75	2	1	40	A4B2	9	7	A3B1	2
11	0.74	2	1	-	A3B1	NT	7	A3B1	NT
12	0.74	2	1	16	A3*B1*	4	35	A3*B1*	2
13	0.72	2	1	16	A3*B1*	4	7	A3*B1*	1
14	0.73	2	1	16	A3*B1*	4	7	A3*B1*	1

d= members of the same family, see Table 4.15

mis-classified as heterozygous null for C4A were borderline. Clearly then, borderline subjects were the main source of error in false-positive assignments to group 1.

One interesting family is indicated by "d" in Tables 4.15, 4.16 and 4.17. Family members 10 (mis-classified, Table 4.15), 2 (borderline, Table 4.17), 8 and 9 (extreme, Table 4.16) all showed the same pattern of lower A or higher B than expected from their genotypes. However, there was no MHC haplotype common to all (Figure 4.2).

4.10. C4A:C4B ratios for various C4 genotypes

C4A:B ratios for various C4 genotypes were evaluated, to determine whether particular C4 genotypes were associated with high or low ratios. Subjects were classified by group according to family studies (Table 4.18). In null and non-null genotypes, the range of ratios was greatest in those where one haplotype was C4A3B1. In all three groups, genotypes carrying C4A6B1 on one haplotype had the lowest mean A:B ratios for their group, although there were only three or four C4A6B1 samples in each genotype group. The ratios for four of the A6B1 subjects were sufficiently low, that they were mis-classified by densitometry as belonging to group 2 or 1, when by family studies, they belonged to groups 3 and 2, respectively (Table 4.15). Thus, C4A6B1 was increased nearly three-fold in mis-classified genotyped subjects (Table 4.22), and the overall trend for A6B1 individuals was towards low C4A/high C4B.

4.11. Total C4 levels for genotyped subjects by single radial immunodiffusion

Total serum C4 levels for 78 genotyped subjects showed that most values observed were within the normal range for serum C4 in healthy Caucasians (0.20-0.50 mg/ml), and all were within the widest reported range (0.06-0.90 mg/ml). With one exception, all values for each category were within mean $\pm$ 2 SD, but the SD for each genotype group was high and the means were not significantly different ($p < 0.07$), (Figure 4.3). Interestingly, eight of ten subjects with C4A6B1 genotypes had lower total

Figure 4.2 **Pedigree for family "D", whose members tend toward less C4A/ more C4B than expected from C4 genotype.** 8 and 9 had extreme A:B ratios; 10 was mis-classified and borderline; 2 was borderline; and 26 * had the lowest total serum C4 by SRID for the heterozygous null C4A group.

A B C D - MHC haplotypes

FAMILY "D"

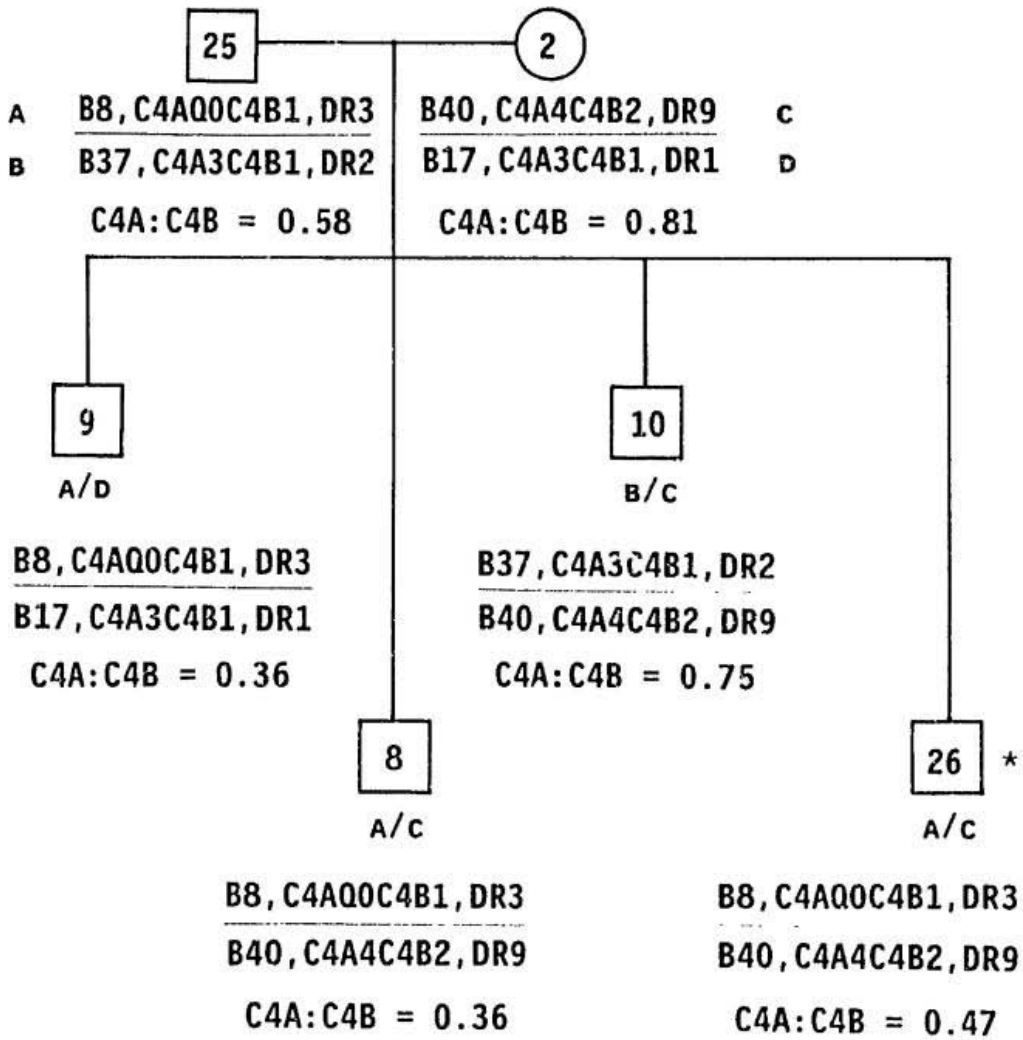

25
2
A B8, C4AQ0C4B1, DR3
B B37, C4A3C4B1, DR2
C4A:C4B = 0.58
B40, C4A4C4B2, DR9 C
B17, C4A3C4B1, DR1 D
C4A:C4B = 0.81
9
A/D
B8, C4AQ0C4B1, DR3
B17, C4A3C4B1, DR1
C4A:C4B = 0.36
10
B/C
B37, C4A3C4B1, DR2
B40, C4A4C4B2, DR9
C4A:C4B = 0.75
8
A/C
B8, C4AQ0C4B1, DR3
B40, C4A4C4B2, DR9
C4A:C4B = 0.36
26 *
A/C
B8, C4AQ0C4B1, DR3
B40, C4A4C4B2, DR9
C4A:C4B = 0.47

Table 4.18

C4A:C4B densitometric ratios for various C4 genotypes

C4 genotype	n[a]	GFS[b]	Mean ratio for group	Mean ratio for C4 genotype	Range of ratios for C4 genotype
		1	0.54		
AQ0B1/A3B1	26			0.61	0.21-1.14
AQ0B1/A3B2	3			0.46	0.39-0.52
AQ0B1/A4B2	7			0.48	0.36-0.59
AQ0B1/A6B1	3			0.42	0.38-0.47
		2	0.97		
A3B1/A3B1	18			0.98	0.49-2.31
A3B1/A4B2	9			1.08	0.75-1.34
A3B1/A6B1	4			0.84	0.58-1.17
A3BQ0/AQ0B1	6			0.96	0.80-1.23
		3	2.01		
A3BQ0/A3B1	9			1.84	0.44-2.85
A3BQ0/A4B2	6			2.14	1.12-2.81
A3BQ0/A6B1	3			1.44	1.01-1.88
Total	94				

[a]genotypes with < 3 samples were not considered

[b]GFS - group by family studies

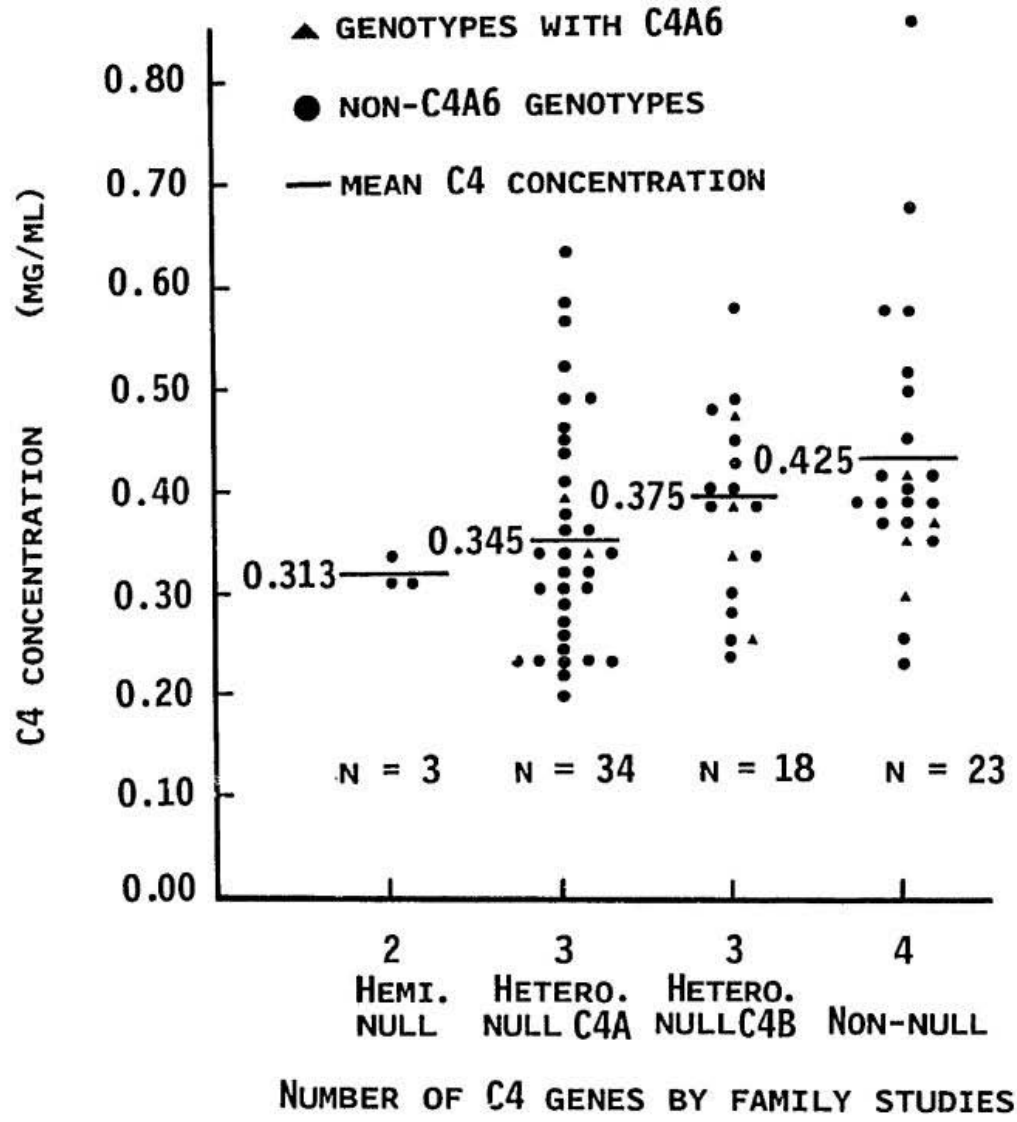


Figure 4.3 **Observed total serum C4 concentrations by SRID for genotyped subjects**

Hemi. = hemizygous

Hetero. = heterozygous

serum C4 than the mean for their corresponding group (Figure 4.3). This was consistent with the observation that A6B1 was over-represented among genotyped subjects in the low C4A/high C4B category.

The C4A:C4B ratios for subjects with the highest and lowest total serum C4 values in each genotype group were evaluated to determine whether particular A:B ratios would correlate with very high or very low total serum C4. For example, an A-null sample with a very high total C4 might show an extremely low A:B ratio, suggesting that the high total C4 was due to extra C4B. (Some of the samples tested by SRID were not analyzed densitometrically. When this was the case for the sample with the highest or lowest SRID value in a group, the A:B ratio for the sample with the second highest or second lowest value was evaluated.) In the heterozygous null C4A group, the A:B ratio for the individual with the highest serum C4 concentration was 0.54, as compared with 0.47 for the individual with the lowest concentration. The lowest sample belonged to family "d" mentioned previously (p. 78), and shown in the figure on p. 79-80. In the heterozygous null B group, the A:B ratio for the second highest sample was 3.30, compared with 2.10 for the lowest sample. Lastly, for the non-null group, the highest sample had an A:B ratio of 1.30, while the second lowest had a ratio of 1.20. These two samples came from siblings who shared identical haplotypes. Thus, there was no correlation between particular A:B ratios and very high or very low total serum C4.

4.12. Densitometric results for phenotyped population members

The considerable range of densitometric ratios for each phenotyped population (Table 4.19), was consistent with the results for genotyped subjects.

Each phenotyped individual was assigned to group 1, 2, or 3 on the basis of A:B ratio, using the ranges derived by discriminant analysis of the genotyped panel. This was done using the SYSTAT MGLH module (pp. 45-46). A subject whose A:B ratio was beyond the lower limit of the range for group 1 of the genotyped panel, or the upper limit of group 3, was assigned a probability of 1 for membership in the appropriate

TABLE 4.19

Range of C4A:C4B ratios in each population tested

Population	n^a	n^b	Range of C4A:C4B ratios Controls	Patients
Newfoundland	108	80	0.18 - 4.64	0.19 - 4.08
St. Louis	122	70	0.25 - 4.83	0.21 - 5.35
Hungarian	24	40	0.37 - 3.15	0.27 - 3.43

n^a number of samples in healthy populations

n^b number of samples in disease populations, i.e. Newfoundland RA, St. Louis SLE, Hungarian SLE

group. For example, the St. Louis SLE subject with a ratio of 5.35 (Table 4.19) was assigned to group 3 (upper limit of range = 3.33) with a p = 1.

Phenotyped population members with extremely low ratios, extremely high ratios, and ratios borderline for group membership were evaluated to determine whether particular C4 phenotypes, or where available, HLA-B,DR phenotypes were associated with unusual C4A:C4B ratios.

Tables 4.20a and 4.20b show C4 phenotypes for population members with extreme ratios. In the high A/low B category, there were seven phenotypes (Table 4.20a). Four had two different C4A genes, but none had two different C4B genes. This suggests that many of these samples carry a null C4B gene. Ten phenotypes contained C4A3A2. A heteroduplicated haplotype C4A3A2BQ0 has been detected by others and is associated with HLA-B35,DR1 (p. 105). These ten individuals probably have three C4A genes and one C4B gene. HLA data were available for five of them. Three had B35,DR1, one had B44/60,DR1 and one had B35,DR2/4. This suggests that densitometry was identifying individuals with duplicated genes. There were HLA data for 22 of the 46 high ratio individuals. Of these, 12 had phenotypes which included C4A3, B44 and DR4, and only one type of C4B gene, suggesting that the haplotype B44, C4A3BQ0 found in the high ratio members of the genotyped panel was also frequent here.

Table 4.20b shows C4 phenotypes for members in the low ratio category. Only one had two different C4A genes, suggesting that the null C4A gene was prevalent. HLA-B and DR data were available for 24 individuals. Of these, 15 had C4B1 and either HLA-B8 o. DR3 or both. This suggests that, as in the genotyped panel, the haplotype B8, C4AQ0B1, DR3 occurred frequently in individuals with very low C4A or very high C4B. Six phenotypes in the low ratio group contained C4A6. This finding was also consistent with data from the genotyped panel.

In the phenotyped populations, all borderline subjects except one belonged to groups 1 or 2. One borderline individual in the St. Louis SLE population belonged to group 3. Table 4.21a shows the borderline population members with marginally high

TABLE 4.20a

C4 phenotypes of population members with extremely high A:B ratios

C4 Phenotype	H_C	H_{SLE}	SL_C	SL_{SLE}	NF_C	NF_{RA}	Total
A3A2B1	1	1	3	2	1	1	9
A3A2B2				1			1
A3B1			11	2	7	9	29
A4A3B2			2		1		3
A4A2B1			1				1
A2B1				2			2
A3B3					1		1
Total							46

Abbreviations: H = Hungarian
SL = St. Louis
NF = Newfoundland
C = Control

TABLE 4.20b

C4 phenotypes of population members with extremely low A:B ratios

C4 Phenotypes	H_{SLE}	SL_C	SL_{SLE}	NF_C	NF_{RA}	Total
A6B1	1	2		1	1	5
A6B2B1		1				1
A3B3B1	1					1
A3B1	2	4	9	7	5	27
A4A2B1		1				1
A5B1		1				1
A2B2B1		1				1
A4B2B1		1		1	2	4
Total						41

ratios. Again, four subjects were C4A3A2B1. This was consistent with the extremely high ratio group. Similarly, three C4A6B1 subjects had borderline low ratios and the borderline St. Louis SLE population member belonging to group 3 was C4A6A3B1 (Table 4.21b). This was consistent with the extreme low ratio group.

4.13. Frequency of C4A6 in phenotyped population members appearing to have low C4A/high C4B by densitometry

The frequency of C4A6 among borderline and extreme population members in the low ratio group compared with their corresponding total populations, was quite striking. C4A6 was from 1.6 to 6 times more frequent in these groups than in the corresponding total populations (Table 4.22). Where haplotype data were available, C4A6 segregated with HLA-B17, DR7. This was consistent with the mis-classified subjects in the genotyped panel in which the frequency of C4A6B1 among subjects in the low ratio group was increased nearly three-fold, relative to the total panel.

4.14. Observed and expected frequencies for groups 1, 2 and 3 in the populations tested

The observed frequencies for groups 1, 2 and 3 were obtained from the densitometry results by counting the number of samples in each group. The expected frequencies for C4A and C4B genotypes were calculated assuming Hardy-Weinberg equilibrium, using the observed frequencies of homozygous null C4A and C4B in each population tested (Table 4.23). (For a discussion of the assumptions upon which this approach is based, see pp. 111-112.) For the Newfoundland and Hungarian populations, the total number in each population is greater than the total reported for densitometry results (Table 4.19), because it includes samples excluded from the densitometric analysis e.g. C4A3B4. In the St. Louis population, C4 null homozygotes had been identified by the 1987 study of Kemp *et al.* Therefore, these samples were not subjected to electrophoresis. Since the total number of samples run did not include any C4 null homozygotes, the

TABLE 4.21a

C4 phenotypes for borderline population members with high A/low B

C4 Phenotypes	SL_C	SI_{-SLE}	NF_C	NF_{RA}	Total
A3B1	4	3	4	2	13
A4B2B1		1			1
A6B1		1			1
A3A2B1			4		4
A3A2B2B1			1		1
A3B3B1			1		1
A3B2B1				1	1
Total					22

TABLE 4.21b

C4 phenotypes for borderline population members with low A/high B

C4 Phenotypes	SL_C	SL_{SLE}	NF_C	NF_{RA}	Total
A3A2B1	2				2
A6A3B1	1	1			2
A6A4B2B1				1	1
A3B1	1	4	5	3	13
A3A2B2B1		1			1
A4B2B1			1		1
Total					20

TABLE 4.22

Frequency of C4A6 in subgroups with low C4A/high C4B by densitometry

Population	Number of C4A6 in subgroup	(%)	Number of C4A6 in total population	(%)
Mis-classified				
genotyped subjects	4/16	(25)	11/119	(9)
Borderline				
St. Louis controls	1/4	(25)	9/128	(7)
St. Louis SLE	1/6	(17)	3/74	(4)
Newfoundland RA	1/4	(25)	4/92	(4)
Extreme				
Hungarian SLE	1/4	(25)	3/42	(7)
St. Louis controls	3/11	(27)	9/128	(7)
Newfoundland controls	1/9	(11)	8/120	(7)
Newfoundland RA	1/8	(12)	4/92	(4)

TABLE 4.23

Frequency of homozygous null C4A and C4B in the populations tested

n		Population	No. of homozygous null C4A (%)		No. of homozygous null C4B (%)	
			Healthy	Patients	Healthy	Patients
120[a]	92[b]	Newfoundland	4 (3.3%)	5 (5.4%)	3 (2.5%)	6 (6.5%)
398[a]	96[b]	St. Louis	6 (1.5%)	10 (10.4%)	15 (3.8%)	2 (2.1%)
38[a]	42[b]	Hungarian	3 (7.9%)	1 (2.4%)	2 (5.3%)	0 (0.0%)
	84*			2 (2.4%)		1 (1.2%)

[a] healthy individuals

[b] patients

* Hungarian SLE population size doubled in order to calculate an expected frequency for homozygous null C4B

frequency of homozygotes in this population was found by counting the number from a list of the total samples supplied. This count agreed with the published values of Kemp *et al.* (1987) for this population.

The expected values were calculated as follows. The frequencies for C4 haplotypes AQ0BX, AXBX and AXBQ0 (with "X" indicating non-null alleles), were represented by fA, fAB and fB, respectively. The double-null haplotype AQ0BQ0 was not included because it is exceedingly rare. None of the 600-plus individuals typed in this study was homozygous for this haplotype, i.e. C4-deficient. All possible haplotype combinations (i.e. genotypes) are shown in Table 4.24. The sum of these genotype frequencies in the equation below, can be taken as 1, given the rarity of the double-null haplotype mentioned above.

$$fA^2 + f(AB)^2 + fB^2 + 2fAfAB + 2fAfB + 2fBfAB = 1$$

AQ0BX / AQ0BX	AXBX / AXBX	AXBQ0 / AXBQ0	AQ0BX / AXBX	AQ0BX / AXBQ0	AXBQ0 / AXBX
0:2	2:2	2:0	1:2	1:1	2:1

Beneath each term in the equation, the corresponding genotype is shown. Below this, the expected C4A:C4B ratio is written. A densitometric ratio cannot be calculated for homozygous null C4A or C4B individuals, represented by 0:2 and 2:0, respectively. Also, the 2:2 and 1:1 groups cannot be distinguished densitometrically, therefore both are represented by the non-null group, 2. That leaves as the other two groups identified densitometrically, group 1 (1:2), and group 3 (2:1). Using the frequencies of homozygous null C4A and null C4B in the St. Louis controls as an example, (Table 4.23), the expected frequency of the 1:2 group was calculated as follows: Six of 398 individuals were homozygous null C4A, so that the frequency of homozygous null A (fA^2) was 6/398 or 0.0151. Thus the frequency of the heterozygous null A haplotype (fA) was calculated as the square root of 0.0151, which was 0.123. Similarly, 15/398 or 0.0377 represented the frequency of homozygous null B (fB^2) in this population. Thus, the frequency of

Table 4.24

Genotype frequencies in a population based on the haplotype frequencies in that population

Haplotype frequencies	fA	fAB	fB
fA	fAfA	fAfAB	fAfB
fAB	fAfAB	fABfAB	fABfB
fB	fAfB	fBfAB	fBfB

f = frequency
A = C4AQ0BX, where X is a non-null C4 allele
AB = C4AXBX
B = C4AXBQ0

the heterozygous null B haplotype (fB) was calculated as the square root of this, which was 0.194. The frequency of non-null haplotypes (fAB) in this population was 1 minus the sum of the null haplotypes, or 1-(fA + fB), which was 0.683. From the formula representing the sum of genotypes shown above, the expected frequency for the 1:2 group was 2fAfAB, or 2(0.123)(0.683) = 0.168. The expected frequencies for the 1:1 and 2:1 groups, calculated as per the formula above, were 0.514 and 0.265, respectively. (Recall that the 1:1 group was equal to the sum of the 1:1 group, 2fAfB, plus the 2:2 group, $f(AB)^2$). So, the total for the three groups analyzed densitometrically was 0.168 + 0.514 + 0.265 or 0.947, with the remaining 0.053 representing homozygous null subjects not analyzed by densitometry. It should be noted that the total of 0.947 included C4A3B4 samples which were, in fact, not included in the observed ratios. When the expected frequency for non-null haplotypes was corrected to account for the C4A3B4 samples, the effect was to lower the expected value for group 2. However, the difference was so small, that uncorrected values were used. The final number of subjects expected in each group was calculated as a fraction of the total for the three groups analyzed. Thus, for the St. Louis controls in group 1, the expected (E) value was 0.168/0.947 = 0.177. Since there were actually 122 subjects analyzed in the St. Louis control population, the expected number of group 1 subjects out of 122 was 0.177 x 122 = 22 (Table 4.25). For all populations, the haplotype frequencies which were used to calculate expected values for groups 1, 2 and 3 are shown in Appendix B (p. 133). It may be noted from the appendix data that although there were no homozygous null C4B individuals in the Hungarian SLE population, approximate E group values were calculated for this population by doubling the number of samples in the population, which gave 84, and assuming that in this number, there would be one homozygous null C4B individual.

Tables 4.25 and 4.26 show observed values by the densitometry-discriminant analysis method and expected values from homozygous null frequencies (Table 4.23). The two populations which most closely matched expected values were the St. Louis SLE ($\chi^2 = 0.25$, $p > 0.80$) and the Hungarian healthy population ($\chi^2 = 0.80$, $p > 0.50$). Of the six populations tested, these two populations had the highest frequencies of

TABLE 4.25

Comparison of observed densitometric frequencies for C4 genotypes with expected frequencies, based on frequency of homozygous nulls

Population (Healthy)	n	Group 1 O	1 E	2 O	2 E	3 O	3 E
Newfoundland	108	34	28	45	56	29	24
St. Louis	122	37	22	51	66	34	34
Hungarian	24	6	8	11	10	7	6

TABLE 4.26

Comparison of observed densitometric frequencies for C4 genotypes with expected frequencies, based on frequency of homozygous nulls

Population (Patients)	n	Group 1 O	1 E	2 O	2 E	3 O	3 E
Newfoundland	80	18	22	43	34	19	24
St. Louis	70	28	27	28	30	14	13
Hungarian	40	17	9	16	24	7	7

homozygous null C4A (Table 4.23). They also bore the closest resemblance to the genotyped reference panel. For example, in the St. Louis SLE population, the expected proportion of subjects in groups 1 and 3 was 39% and 13% respectively, compared with 44% and 19% in the genotyped panel.

In contrast to the results for the St. Louis SLE and the Hungarian healthy populations, the observed and expected distributions for the St. Louis control population and the Hungarian SLE population were significantly different ($\chi^2 = 13.6$, $p < 0.005$ and $\chi^2 = 9.7$, $p < 0.01$, respectively). In both populations, more samples were assigned to group 1 and fewer to group 2 than expected, assuming Hardy-Weinberg equilibrium. Of the six populations tested, these two had the lowest frequencies of null C4A alleles (Table 4.23). They also bore the least resemblance to the genotyped panel. For example, in the St. Louis control population the proportion of subjects in groups 1 and 3 was 18% and 28%, respectively, compared with 44% and 19% in the genotyped panel. The results for the St. Louis control population and the Hungarian SLE population suggested that: (1) as in the genotyped reference panel, the densitometric method overestimated group 1 heterozygous null C4A, and (2) the accuracy of the method was greatly affected by the frequency of null C4A in the populations tested. It appeared to be most accurate for the populations with high null C4A frequencies and least accurate for those in which the frequency of null C4A was low.

The Newfoundland control population was the population with the lowest frequency of homozygous null C4A individuals in which observed and expected frequencies were not significantly different ($\chi^2 = 4.5$, $p > 0.10$). This suggested that, for a population in Hardy-Weinberg equilibrium, the densitometry-discriminant analysis method would produce reliable results in a population with a minimum null C4A frequency of approximately 3% (Table 4.23).

When expected values for the diseased and control populations were compared, only the St. Louis SLE and their controls were significantly different ($p < 0.001$), with the largest component of $\chi 2$ belonging to group 1.

In contrast, when observed values were compared for the diseased and control populations, the two St. Louis populations were no longer significantly different ($p > 0.10$). However, the Hungarian SLE and their controls were significantly different ($p < 0.05$), again, with the largest component of χ^2 belonging to group 1.

In summary, these results suggest that gene frequencies in a population may affect, to some degree, the accuracy of densitometry for assigning null genes to heterozygotes. This implies that the method may only be reliable when applied to populations having a certain profile.

Chapter 5

DISCUSSION

Because no systematic studies have been done on the accuracy of densitometry as a predictor of genotype, this discussion will focus on the factors which could contribute to error in using the method for this purpose.

The results of this study suggest that subjects can be "mis-classified" by the densitometry-discriminant analysis method for the following reasons:

(1) Experimental design

(2) Inaccurate pedigrees

(3) Hidden duplications

(4) Altered synthesis/catabolism of C4

(5) Sample quality

5.1. Effect of experimental design on mis-classification

5.1.1. Discriminant analysis

Discriminant analysis maximizes the accuracy of prediction of group membership, and thus, was a logical choice for this study. However, since the cutting scores for each group are based on the mean of each group (p. 44), this type of analysis is affected by the range of the ratios which contribute to the mean of each group. For example, excluding the group 3 subject with the lowest A:B ratio, (Table 4.9), would have raised the mean for group 3 and the cutting scores for each group. As a result,

there would have been fewer false-positive assignments to each group, and in particular, fewer subjects would have been mis-classified as heterozygous null for C4A. In this study, degraded samples such as the one referred to above were excluded on a subjective basis. Outliers could have been excluded from each group by a statistical means, before subjecting the data to discriminant analysis. (For a discussion of the statistical treatment of outliers, see the National Bureau of Standards' *Experimental Statistics*, 1966). Eliminating outliers would decrease the overlap between groups, and this would reduce the number of borderline subjects. These individuals are most difficult to classify by densitometry, and they are the main source of error in false-positive assignments to group 1. Therefore, reducing the number of borderline subjects would also reduce the number of subjects mis-classified as heterozygous null for C4A. However, while excluding outliers would have made the analysis look more accurate, those outliers which were not the result of sample degradation represented a genuine result i.e. extreme ratios, and for this reason should not have been excluded.

5.1.2. Composition of the genotyped panel

The results obtained using densitometry plus discriminant analysis, may also be affected by the composition of the reference (genotyped) panel. In this study, null C4A heterozygotes made up the largest group, null C4B heterozygotes made up the smallest group, and this may have shifted the reference ranges. This could be tested by using equal numbers in each group or using proportions which more closely resemble the expected frequencies. It is noteworthy that the method worked best for the phenotyped populations whose expected frequencies were closest to the composition of the reference panel (p.95, 98). If this is not coincidence, it suggests that a reference panel would have to be tailored for each unknown population, and thus greatly limits the usefulness of the method.

5.2. Limitations of densitometry

There are some samples which cannot be effectively analyzed by densitometry. These are those with the allotype C4A3B4; samples from hemizygous null individuals with AQ0 and BQ0, and those from individuals with hidden duplications. C4B4 occurs with a frequency of close to 1% in Caucasians (Baur *et al.* 1984) and hemizygous null individuals occur with a frequency of about 2% (Welch *et al.* 1985). The frequency of hidden duplications is unknown, yet despite this, relative A:B ratios are still used by many investigators to assign null alleles.

The null A and B alleles in hemizygous null individuals cannot be detected densitometrically, because these individuals are likely to have A:B ratios of approximately one. Single null alleles can also remain undetected by densitometry if they occur with hidden duplications. For example, an individual with the genotype A3A3B1/AQ0B1 would have the phenotype A3B1 and a densitometric ratio of approximately 1. The null allele occurring with the duplication would only be detected in a study of an informative family, as in the following example.

a	A3A3,B1	x	AQ0B1	c
b	A3,B1		AQ0B1	d

(Informative family)

Possible genotypes of off-spring:

	(axc) A3A3B1/AQ0B1	A3A3B1/AQ0B1 (axd)
(C4A:B ratio)	1:1	1:1

AQ0 on haplotypes c and d would not be detected if individuals were not part of a family study

	(bxc) A3B1/AQ0B1	A3B1/AQ0B1 (bxd)
(C4A:B ratio)	1:2	1:2

AQ0 is detected

5.3. Inaccurate pedigrees

When one method is being compared to another standard method, the ideal would be a thoroughly trustworthy ("gold") standard. In practice, this is not possible. The gold standard chosen for this study (group by family studies), while admittedly imperfect, was considered among the best available. Obviously, a method can perform no better than the standard against which it is judged. There were two potential sources of error in the gold standard used in this study. First, approximately 18% of the genotyped controls were not secure. That is, a few families were not informative for all C4 alleles, so known HLA-associations were used to assign some haplotypes (pp. 32-33).

For the purpose of this study, it would have been best if these subjects had not been included. However, it was not discovered until this thesis was being written, that such individuals had been included in the genotyped panel. It was decided to keep them as part of the genotyped panel, because most investigators make use of HLA data in interpreting pedigrees, therefore including them in this study, reflects the method in common use. The mis-classified group, (Table 4.15), suggests that some such individuals were being identified by densitometry. For example, if subject 17 had one null C4A allele, this could explain the apparent low C4A/high C4B in this person. The large number of individuals typed (*) in the mis-classified samples, suggests that there is an error associated with the practice of genotyping using HLA-association data, and that densitometry may be more accurate than it appeared to be in this study.

5.4. Interpreting densitometric results for the genotyped panel

As expected, the mean ratio for group 1 approached 0.5, for group 2 approached 1.0, and for group 3 approached 2.0. Although the analysis of variance showed a significant group effect, the range of ratios for each group was large, and there was considerable overlap between groups. Thus, for many samples, A:B ratio alone could not be used to predict genotype or to assign null A or B genes.

The conclusions regarding the densitometry-discriminant analysis method, based on the genotyped panel are:

(1) The overall accuracy of densitometric ratio as a predictor of genotype is, in this analysis, 78% (p. 63).

(2) The most common mis-classification error is falsely identifying subjects as heterozygous null for C4A (p. 71).

(3) The overall trend in mis-classification is shifting subjects from group 3 to group 2, and from group 2 to group 1 (p.72).

The figure arrived at for overall accuracy is an overestimate because, although it was not the case in this study, within and between-run variability could contribute to mis-classification. Observations (2) and (3) suggest that individuals tend to have lower C4A or higher C4B than expected from their family genotype. This shift may be an artifact caused by machine error or the discriminant analysis, or it may represent a true difference in C4A and C4B levels. The former possibility has already been discussed. If individuals do tend to have lower A/ higher B than expected, then this suggests that C4A and C4B are not necessarily produced/catabolized at the same rate, or that some individuals carry hidden duplications of C4B.

Recently, a group used two different genetic models to compare the results of C4 genotyping by electrophoretic gels with genotyping by family haplotype and disequilibrium data (Klitz, Borot and Thomson, 1988). The additive model assumes an individual heterozygous null for C4A can be typed directly from a gel, while the recessive model does not. Their results demonstrated that mis-classification can be due to both missing null alleles in heterozygotes and to falsely assigning null alleles. Interestingly, they also found that C4B null had a higher assignment error rate than C4A null, the same result shown in this study by the false-negative assignments for B-null (pp. 70-71). They concluded that the additive model is not correct, and that even with pedigree data, some mating types are impossible to type accurately.

5.4.1. Hidden duplications

The following hypotheses have been made regarding the origin of hidden duplications. Unequal crossing-over or gene conversion could yield two identical alleles at the C4 loci, (e.g. C4A3A3,B1 or C4A3A3, no C4B), or C4 alleles of the same isotype but different allotype (e.g. C4A3A2,BQ0) on one chromosome (Yu and Campbell, 1987; Palsdottir *et al.* 1987b). The first possibility is referred to as homoduplication, the second, as heteroduplication (Yu and Campbell, 1987).

Heteroduplications are well known. For example, the C4A3,A2 duplication is observed with C4BQ0, most commonly on the haplotype HLA-B35, DR1 (Uring-Lambert *et al.* 1984; Rittner *et al.* 1984b; Raum *et al.* 1984). C4B1,B2 has been observed with C4A2 on HLA-B7,DR3 or DR5 and HLA-B14,DR1, 3 or 6 (Raum *et al.* 1984; Uring-Lambert *et al.* 1984). Heteroduplications can be detected by family studies and conventional C4 typing because the alleles have a different charge. In contrast, homoduplications cannot be detected by family studies. Rather, they must be inferred from relative band density on typing gels or from RFLP patterns. A limited number of studies have suggested homoduplications as follows:

(1) By quantitative increases in C4A bands on C4 typing gels -

C4A3 with C4B1 or BQ0 and various (unspecified) HLA-types (Uring-Lambert *et al.* 1984)

(2) By RFLP analysis -

(i) C4A3 with BQ0 on HLA-B44,DR6 (Yu and Campbell, 1987) and on HLA-B44,DR4 (Truedsson *et al.* 1987) i.e. C4A3 is substituted for C4B, because the B locus has been converted to express the A locus product, giving A3A3BQ0.

(ii) C4B1 with AQ0 on HLA-B8,DR3 (Truedsson *et al.* 1987) and on HLA-B17,DR6 (Schneider *et al.* 1986), resulting in AQ0B1B1

(iii) C4B1 with A3 on HLA-B44,DR7 (Schneider *et al.* 1986) and with HLA-B44 (Palsdottir *et al.* 1987a), resulting in A3B1B1.

Thus, homoduplications have been inferred on null and non-null haplotypes. In particular, HLA-B44 may be associated with homoduplication of either C4A or C4B.

In this study, more than half of the mis-classified genotyped panel subjects with a higher A:B ratio than expected from genotype carried the haplotype HLA-B44, C4A3B1 (Table 4.15). Therefore, some of these subjects may have homoduplicated C4A3. For example, subjects 3 and 4 may carry C4A3A3B1 on the B44,DR8 haplotype. Similarly, in the low C4A/high C4B category, subject 15 may have homoduplicated C4B1 on the B44,DR2 haplotype.

Homoduplications may also be inferred from some extreme A:B ratios observed in this study. For example, all of the subjects with extremely high A:B ratios (1-5, Table 4.16) carried the haplotype HLA-B44, C4A3BQ0. Most non-deleted C4B null alleles segregate with HLA-B44 (p. 20). Therefore, although the C4B gene in these extreme subjects is probably not deleted, it is also not expressed as the product of the B locus. This could be because it is a non-functional/defective gene, or it may have been converted to express the A locus product. Because these ratios indicate high C4A relative to C4B, the latter suggestion seems more likely.

While a high C4A:C4B ratio is associated with HLA-B44, C4A3B1 or C4A3BQ0, a low ratio is associated with C4A6B1 (9, 11, 19, 20, Table 4.15). Among mis-classified subjects, C4A6B1 appeared only in the low C4A/high C4B category, and was increased nearly three-fold in this group as compared with the total genotyped panel. Also, when densitometric ratios for a variety of C4 genotypes were compared, those genotypes which included C4A6B1 had mean C4A:C4B ratios consistently below the mean for the corresponding group, based on gene number by family studies (Table 4.18). These results suggest that some A6B1 chromosomes may have homoduplicated C4B1.

5.5. Total serum C4 concentrations

5.5.1. For low C4A:C4B ratio samples

The alternatives to gene duplication for apparent increases or decreases in C4A or C4B protein, are altered synthesis of C4 or different catabolism of C4A and C4B. That is, C4A and C4B may be produced/used up at different rates, and different variants of C4A or C4B may be produced/used up at different rates. Densitometric ratios cannot provide direct proof for either a synthesis/catabolism or duplication mechanism. C4A:B ratios for A6B1 subjects suggested duplication of B1 on some chromosomes. If this were so, then one would expect that the total serum C4 for some individuals whose genotype included C4A6, would be increased. Therefore, total C4 concentrations were tested by SRID in order to evaluate this hypothesis. The SRID results for C4A6 did not support the hypothesis of duplicated B1 (Figure 4.3). In the heterozygous null C4B group and in the non-null group, most C4A6 individuals fell below the mean total serum C4 concentration. This suggests that in some subjects bearing C4A6, there may be a reduction in synthesis of C4A6, or an increase in its catabolism, rather than a duplication of C4B1 on the A6B1 chromosome. It also suggests that not all C4A variants are produced/catabolized at the same rate.

C4A6 codes for a hemolytically inactive product. It is the only C4 variant identified to date in which allotypic variation affects function. This C4 variant is hemolytically inactive because it contains a mutation which destroys the site on its β-chain by which C3b must bind to form an active C5 convertase (Kinoshita *et al.* 1988). Whether and how this defect in function might be related to reduced C4A6 synthesis or increased catabolism is open to speculation.

There is some evidence to support altered C4 synthesis as the basis of C4 deficiency. Muir *et al.* (1984) and Wisniewski *et al.* (1987) have identified a C4-deficiency caused by hyposynthesis of C4 regulated by a gene not linked to the MHC. In some cases, rather than hyposynthesis of a C4 product there may be no synthesis, due to a completely

non-functional gene. This has been shown to be the case for some homozygous null individuals in which the null gene on one chromosome is not deleted, but clearly is non-functional (p. 25) and in completely C4 deficient individuals who show no gene deletions (Hauptmann *et al.* 1987; Goldstein *et al.* 1988).

5.5.2. For high C4A:C4B ratio samples

Extremely high C4A:B ratios in some genotyped subjects suggested a possible homoduplication of C4A3 on the B44 chromosome, or high production of A3. SRID results showed that 3/5 of these individuals had higher total serum C4 than the mean for their group, suggesting either duplication or high production of C4A3.

SRID was used in this study because it is a standard method for measuring total serum C4 concentrations, and the equipment for the more sophisticated alternative of rate nephelometry was not available. However, there are limitations to SRID, as discussed below, and these must be kept in mind when evaluating the results discussed in the preceding two sections.

Total serum C4 concentrations for 78 genotyped subjects, as measured by SRID, showed that there was not a significant association between gene number by family studies and total serum C4 (Figure 4.3). This is in agreement with numerous other reports, which show that serum C4 concentrations are not strictly correlated with the number of C4 genes (Olaisen *et al.* 1980; Awdeh *et al.* 1981; Rittner *et al.* 1984b; Welch *et al.* 1985; Sjöholm *et al.* 1985; Uko *et al.* 1986). If one assumes that the pedigrees are accurate, then these results do not support the hypothesis of gene-dose effect.

In healthy individuals there is a considerable range of serum C4 concentrations. This indicates that individuals vary in the total amount of C4 they express. It follows that individuals could show the same relative amounts of C4A and C4B (e.g. 1:1), but very different total C4 concentrations. The A:B ratios for extreme total serum C4

concentrations by SRID in this study support this reasoning. When C4A:B ratios were compared for subjects representing the two extremes of serum C4 concentrations for each group, the ratios were, in nearly every case, virtually identical. For example, two siblings shared identical haplotypes and nearly identical A:B ratios, but one had the highest C4 level in a group, while the other had the lowest (p. 83). Furthermore, combining the densitometric and SRID results for the two subjects with nearly identical A:B ratios (about 1:1) but quite different total C4 concentrations, suggests that in these subjects the relative expression of A and B in each individual was the same although the total expression was different.

From the SRID results in this study, it was not possible to determine if, in some cases, the C4B gene expresses more protein than the C4A gene, or vice versa. One way to determine whether C4A expression differs from C4B expression using SRID, would be to compare serum C4 concentrations for homozygous null (C4A or C4B) individuals.

Although C4A:B ratios cannot conclusively show whether C4A expression differs from C4B expression, the densitometric results in this study suggest that C4B may, in general, be more highly expressed than C4A. This is supported by the following observations. First, two thirds of mis-classified subjects appear to have low A/high B. Second, among extreme subjects in the low A/high B category, half of the haplotypes are HLA-B8, C4AQ0C4B1, DR3. RFLP studies show that most deleted C4A null genes are linked to this haplotype (p. 20). Such deletions would explain low ratios approaching 0.5. Unusually low ratios suggest that C4B may be more highly expressed than usual on this haplotype although it is also possible that C4B may be duplicated on this haplotype (p. 105). Another possibility is that the non-B8,DR3 ("other") haplotype contains the highly expressed or duplicated C4B. However, the low ratio samples in the genotyped panel (Table 4.16) had four different "other" haplotypes, suggesting that this possibility is less likely. There is some evidence to support the possibility that C4B is produced at higher concentrations than C4A. Using an ELISA method to measure individual levels of C4A and C4B, Holme *et al.* (1988) found that there is a poor

correlation between C4A and C4B concentrations, and that some individuals have disproportionately high C4B levels in comparison to C4A. In a study relating C4A and B levels to age in a large group of healthy individuals, they also found that mean levels of C4B increase by approximately 12% per decade, while for C4A the increase is 3%. Also, Campbell and his co-workers have observed that some genotypes, regardless of gene number, tend to have more C4B than C4A (D. Campbell, personal communication). These independent observations suggest that the inference in this study that C4B may be expressed differently than C4A, is unlikely to be due to a systematic error in the technique used.

Clearly, better methods are needed to resolve the synthesis/catabolism and duplication possibilities.

5.6. Phenotyped population members

5.6.1. Evaluation of C4 phenotype, HLA and C4A:C4B ratio data

Evaluation of C4 phenotype, HLA and C4A:C4B ratio data for the populations, showed the same trends observed for the genotyped panel:

(1) A low C4A:C4B ratio is associated with C4A6B1 (4/16 low A:B samples in the genotyped panel and 9/46 low A:B samples in the population study, Table 4.22). Samples from the genotyped panel tested by SRID had low or low normal total serum C4. This suggests that either the chromosome which carries HLA-B17, C4A6 C4B1, DR7 has an A6 gene which is expressed at low levels, or that the antibody detection of A6 may be reduced due to a deleted epitope.

(2) An extremely high C4A:C4B ratio is associated with C4A3 C4BQ0 and HLA-B44 [5/5 high A:B samples in the genotyped panel (Table 4.15) and 12/22 tested in the population study]. These data suggest that the C4A3 on a chromosome containing HLA-B44, C4A3 C4BQ0 is expressed at high levels or, alternatively, that the C4B null gene on this chromosome has been converted to a second A3.

5.6.2. Distribution of null and non-null genotypes

The following is a summary of the main observations for the phenotyped populations:

(1) The populations which most closely matched expected distributions were the St. Louis SLE and the Hungarian controls. Of the populations tested, these two had the highest frequencies of null A homozygotes (p. 95, 98).

(2) Two populations, the St. Louis controls and the Hungarian SLE differed significantly from expected distributions. These populations had the lowest frequencies of null A homozygotes (p. 98).

(3) The greatest shift in the distribution of subjects by densitometry was from group 2 (no nulls) to group 1 (one null A). This shift was greatest in the St. Louis healthy and the Hungarian SLE populations, and is reflected in the component of χ^2 for group 1 (pp. 98-99).

(4) The accuracy of the densitometry-discriminant analysis method of assigning null genes to heterozygotes seems to depend to some extent, on the gene frequencies of the population tested.

These conclusions are based on the assumptions that the populations are in Hardy-Weinberg equilibrium or that estimates of heterozygotes and non-null homozygotes from observed frequencies of null homozygotes are reliable. This may not be true for the Hungarian populations because both are small and therefore likely to be subject to sampling error. The St. Louis populations on the other hand, are moderately large and have been extensively studied by another method (Kemp *et al.* 1987). That study has shown that although the two populations differ significantly with respect to null C4A

homozygotes, the observed frequency of null C4A heterozygotes in each is approximately equal to that predicted assuming Hardy-Weinberg equilibrium.

The results for the two populations which most closely matched expected distributions suggest that the more prevalent or frequent the C4A null allele is in the population tested (St. Louis SLE and Hungarian controls, Table 4.23), the more likely it is to be correctly identified, and the less false-positive assignments will contribute to mis-classification in group 1. In other words, if densitometry is applied to a group of individuals with a low likelihood of having AQ0 (St. Louis controls and Hungarian SLE, Table 4.23), subjects identified as heterozygous null for C4A will be largely false-positives. The results of this study suggest that in order to obtain a reliable estimate of null and non-null genotype distribution in a population using densitometry, the frequency of homozygous null A in the population should be approximately 3% (p. 98).

The St. Louis population best illustrates the limitations of using densitometry to detect null C4A alleles in populations. The same SLE population used in this study, was subjected to DNA analysis by Kemp *et al.* (1987). They found that 34.5% of the SLE patients were heterozygous null for C4A due to a gene deletion. In this study, the observed frequency in this group was 28/70 or 40%, suggesting that about 6% of the SLE patients were false-positive heterozygous null for C4A (Table 4.26). Because RFLP analysis identifies only C4A gene deletions, the 6% false-positive AQ0 in this population may include true positives that have C4A genes which, for reasons other than deletion, are not expressed. As the frequency of homozygous null A in this SLE population was high, and the frequency of homozygous null B was low, the distribution of genotypes obtained by densitometry was nearly accurate. Kemp and co-workers found 12.5% of their controls had the heterozygous C4A deletion, compared to 37/122 or 30% identified as heterozygous null A in this study (Table 4.25). This suggests an 18% false-positive rate in this group, although a few of these may be true positives with non-deleted, non-expressed genes. Because the frequency of homozygous null C4A in this control

population was low, a high number of false-positive assignments were made to group 1, resulting in an inaccurate distribution of heterozygous null A and non-null genotypes. RFLP analysis shows that heterozygosity for C4AQ0 is significantly increased among St. Louis SLE patients. However, the results of this study show that densitometry could not detect this difference due to the method's tendency to falsely identify subjects as heterozygous null for C4A. This effect is most pronounced in populations like the St. Louis controls in which the frequency of C4AQ0 is low. In such a population, differences in genotype frequencies shown by other methods may be inapparent using densitometry.

While the effect of shifting subjects from group 2 to 1 by densitometry is shown best by the results for the St. Louis control population, the distribution of subjects in groups 1-3 for the Newfoundland RA's illustrates the other way in which the method shifts subjects. The nine extra observed subjects in group 2 (Table 4.26) come from groups 1 and 3, in nearly equal numbers. By the same reasoning used previously, if densitometry is applied to a group of subjects with a "high" likelihood of having C4BQ0, a large proportion of those identified as lacking the allele will be false-negatives. The frequency of homozygous null B was highest among Newfoundland RA's (Table 4.23), thus many of the subjects shifted from group 3 to 2 are false-negative for BQ0. The frequency of homozygous AQ0 in this population was nearly equal to that of BQO. Thus, few false-positive assignments to group 1 were made, and as mentioned above, the overall shift in this population was toward group 2. Because the distribution of genotypes in the Newfoundland RA population was not significantly different from its control population, this suggests that C4B null is not significantly increased in this RA population.

If the homozygous null A frequencies in the Hungarian controls and SLE patients examined here are representative of the true frequencies in the entire Hungarian healthy and SLE populations, then the results suggest that healthy Hungarians have a higher frequency of null A genes than the other healthy populations considered in this study, and SLE in Hungarians may not be associated with a significant increase in hetero-

zygosity for null C4A. While there was no significant difference between the Hungarian populations when expected distributions were compared (p. 95, 98), there was when observed values were compared (p. 99). However, these results should be interpreted bearing in mind the following points. (1) Given the small sample size, the frequency of homozygous null C4A and C4B found in this study may not represent the true frequency in the larger Hungarian population. (2) The significant difference for observed values may not be genuine, but rather a reflection of the way in which the method shifts samples towards group 1.

Since the expected distribution for the three groups in the Hungarian control population was not significantly different from the observed, this suggests that densitometry may be useful for predicting the frequency of non-null and heterozygous null B genotypes in small populations. The densitometry-discriminant analysis method could also be used for populations not in Hardy-Weinberg equilibrium.

Many studies have based null C4A/disease associations in population studies using A:B ratios to assign null heterozygotes. The analysis in this study suggests that this is an unreliable way to compare two populations whose null C4A frequencies are hypothesized to vary widely. If there is a tendency to have more B or less A than a simple gene-dose effect would predict, then, in populations of low null A frequency (control populations, for example), the proportion of apparent null A heterozygotes will be artificially high, and the results spurious.

The repeated occurrence of the same haplotypes with low ratios -- B17, C4A6B1, DR7-- and others with high ratios -- B44, C4A3BQ0 -- lends support for the possibility that different C4A or C4B genes may be expressed at different rates and/or that some haplotypes contain hidden 'extra' or converted genes.

In summary, the aim of this study was to evaluate the usefulness of densitometry as a method for detecting null C4 alleles in populations of unrelated individuals. In doing this, a number of questions were raised (p. 29). The conclusions of the study regarding these questions are:

(1) The ratio of C4A to C4B in an individual's serum is an inaccurate predictor of his/her genotype. The overall accuracy is, at best, approximately 80% and may differ for different genotypes. An error rate of 20% is clearly too high for useful genotyping.

(2) Using densitometric measures of C4A:C4B ratio to estimate population frequencies of carriers of null C4 genes tends to overestimate the frequency of carriers of null C4A. This seems to be particularly pronounced in populations in which the actual frequency of the null C4A gene is low.

This limits the usefulness of the method for comparing distributions of genotypes in populations for which gene frequencies are likely to be very different, as for example, in the diseased and healthy populations compared in this study. The results of the present study suggest that disease association studies based on densitometric analyses should by interpreted with caution. For example, one group (Christiansen *et al.* 1983) has used a similar densitometric method and reports good correlation between C4A:C4B ratio and genotype from family studies. However, in a survey of 102 healthy controls and 90 diabetics (Uko *et al.* 1986), they were able to determine the number of null alleles by A:B ratio in only 74 of the controls and in 60 of the 90 diabetics. Two of the 28 control samples were excluded due to overlapping bands. The remaining 26 were termed "indeterminate" which was defined as the presence of only one allele at each locus and a densitometric ratio of 1. Presumably, these 25% represent hemizygous-null individuals, but this frequency is far in excess of that reported (approximately 2%) for these individuals. More likely, this proportion includes individuals heterozygous null for C4A or C4B, who nonetheless have a densitometric ratio of 1, and those who are homozygous non-null. Therefore, this group was able to predict genotype by densitometry for, at best, 75% of their control population. Such findings reinforce the conclusion of this study that densitometry is not practical for determining genotype distributions in various populations.

(3) The 'gene dose effect' is based on the assumptions that all C4A and C4B genes are expressed equally, and that no individuals carry undetected, duplicated A or B genes. The results of this analysis suggest that a sizeable number of individuals may have more C4B than C4A, and that this may not necessarily be related to genotype *per se*. In addition, the data suggest that the MHC haplotype HLA-B17, C4A6C4B1, DR7 may carry a C4A6 that is expressed at low levels, and the haplotype HLA-B44, C4A3C4BQ0 may carry two copies of C4A3 or a single A3 that is expressed at high levels.

Because densitometry is useful for analyzing large numbers of samples relatively quickly, this study concludes that the method's best use would be in sifting through large populations to identify potentially interesting subjects for study by more sophisticated methods.

5.7. Future aims

It is clear that better methods are needed to investigate the relationship between true gene number and individual serum C4A and C4B levels. Such methods could be used to further investigate the questions raised in this study concerning the level of C4B expression relative to C4A in an individual; the possibility that different C4A and C4B variants are produced/catabolized at different rates; and that some of the ratios observed are the result of hidden duplications.

One method which will provide a reliable value for the number of C4 genes in a given phenotype, has just been developed (Dunham *et al.* 1989). Briefly, DNA from peripheral blood mononuclear cells is digested with particular restriction enzymes which cut infrequently on either side of the C4A-C4B gene complex. This generates large DNA fragments which are separated by pulsed-field gel electrophoresis, transferred to membranes and hybridized with genomic and cDNA probes specific for the complement gene cluster. The size of the fragments hybridizing with these probes is directly

related to the number and length of the C4 genes present in the DNA samples analyzed. In this way, both the number and size of C4A and B genes present on a chromosome are revealed. This method may provide a means for distinguishing deleted from non-deleted null alleles, and may prove especially useful for studying persons in whom homoexpression is suspected.

Individual levels of serum C4A and C4B can now be measured using an enzyme-linked-immunosorbent-assay (ELISA), using C4A and C4B-specific monoclonal antibodies recently produced in this laboratory.

One way to proceed further using the results of this study, would be to select from informative families, samples with extremely high and low A:B ratios, those that were mis-classified and those with the C4A6B1 genotype and determine, as far as possible, the true gene number in these samples using the direct DNA analysis method described above. Serum C4A and C4B levels for these samples could be measured using the monoclonal antibody technique. Taken together, results from direct gene counting, direct serum C4 quantitation and family studies should make it possible to elucidate the relationship between serum C4 concentrations and C4 genes, and this will be directly important in disease association studies. In SLE for example, deleted C4A genes occur with increased frequency, and it has been suggested that decreased serum C4A concentration in individuals with one or two deleted C4A genes could result in defective processing of immune complexes (p. 25). One assumption of this hypothesis is that serum C4A concentration and C4A gene number are directly related. Clearly, investigators will be aided best in their efforts to determine the nature of the relationship between null alleles, serum C4 concentrations and autoimmune diseases such as SLE, by using the most accurate methods currently available, in order to dissect this relationship.

REFERENCES

Agnello, V. Complement deficiency and systemic lupus erythematosus. In: *SLE.* Lahita, R.G., ed. John Wiley & Sons, New York pp. 565-589, 1987.

Allen, F.H., Jr. Linkage of HL-A and GBG. *Vox. Sang.* 27: 382-384, 1974.

Alper, C.A., Boenisch, T., Watson, L. Genetic polymorphism in human glycine-rich beta-glycoprotein. *J. Exp. Med.* 135: 68-80, 1972.

Alper, C.A., Johnson, A.M. Immunofixation electrophoresis: a technique for the study of protein polymorphism. *Vox. Sang.* 17: 445-452, 1969.

Averill, B.K., Bernal, J.E. Genetic and environmental influences on serum levels of human C4. *Ann. Hum. Biol.* 11: 149-156, 1984.

Awdeh, Z.L., Alper, C.A. Inherited structural polymorphism of the fourth component of human complement. *Proc. Natl. Acad. Sci. USA* 77: 3576-3580, 1980.

Awdeh, Z.L., Raum, D., Alper, C.A. Genetic polymorphism of human complement C4 and detection of heterozygotes. *Nature* 282: 205-207, 1979.

Awdeh, Z.L., Ochs, H.D., Alper, C.A. Genetic analysis of C4 deficiency. *J. Clin. Invest.* 67: 260-263, 1981.

Awdeh, Z.L., Raum, D., Yunis, E.J., Alper, C.A. Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. *Proc. Natl. Acad. Sci. USA* 80: 259-263, 1983.

Ballow, M., McLean, R.H., Einarson, M., Martin, S., Yunis, E.J., Dupont, B., O'Neill, G.J. Hereditary C4 deficiency. Genetic studies and linkage to HLA. *Transplant. Proc.* 11: 1710-1712, 1979.

Batchelor, J.R., Fielder, A.H.L., Walport, M.J., David, J., Lord, D.K., Davey, N., Dodi, I.A., Malasit, P., Wanachiwanawin, W., Bernstein, R., Mackworth-Young, C., Isenberg, D. Family study of the major histocompatibility complex in HLA DR3 negative patients with systemic lupus erythematosus. *Clin. Exp. Immunol.* 70: 364-371, 1987.

Baur, M.P., Neugebauer, M., Deppe, H., Sigmund, M., Luton, T., Mayr, W.R., Albert, E.D. In: *Histocompatibility Testing 1984.* Albert, E.D., Baur, M.P., Mayr, W.R., eds. Springer-Verlag, Berlin, pp. 333-341, 1984.

Belt, K.T., Carroll, M.C., Porter, R.R. The structural basis of the multiple forms of human complement component. *Cell* 36: 907-914, 1984.

Belt, K.T., Yu, Y., Carroll, M.C., Porter, R.R. Polymorphism of human complement component C4. *Immunogenetics* 21: 173-180, 1985.

Breedveld, F.C., Lafeber, G., Dreves, G., Class, F., Cats, A. Felty's syndrome: autoimmune neutropenia or immune-complex mediated disease. *Rheumatol. Int.* 5: 253-258, 1985.

Campbell, R.D., Gagnon, J., Porter, R.R. Amino acid sequence around the thiol and reactive acyl groups of human complement component C4. *Biochem. J.* 199: 359-370, 1981.

Campbell, R.D., Law, S.K.A., Reid, K.B.M., Sim, R.B. Structure, organization and regulation of the complement genes. *Ann. Rev. Immunol.* 6: 161-195, 1988.

Campbell, R.D., Bentley, D.R., Morley, B.J. The factor B and C2 genes. *Phil. Trans. R. Soc. Lond. B.* 306: 367-378, 1984.

Carroll, M.C., Capra, J.D. Studies on murine Ss protein: demonstration that the Ss protein is functionally fourth component of complement. *Proc. Natl. Acad. Sci. USA* 75: 2424-2429, 1978.

Carroll, M.C., Belt, K.T., Palsdottir, A., Porter, R.R. Structure and organization of the human C4 gene. *Phil. Trans. R. Soc. Lond. B.* 306: 379-388, 1984a.

Carroll, M.C., Campbell, R.D., Bentley, D.R., Porter, R.R. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature* 307: 237-241, 1984b.

Carroll, M.C. Katzman, P., Alicot, E.M., Koller, B.H., Geraghty, D.E., Orr, H.T., Strominger, J.L., Spies, T. Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. *Proc. Natl. Acad. Sci. USA* 84: 8535-8539, 1987a.

Carroll, M.C., Alper, C.A. Polymorphism and molecular genetics of human C4. *Br. Med. Bull.* 43: 50-65, 1987b.

Carroll, M.C. Molecular genetics of the fourth component of human complement. *Federation Proc.* 46: 2457-2462, 1987c.

Carroll, M.C., Belt, K.T., Palsdottir, A., Yu, Y. Molecular genetics of the fourth component of human complement and steroid 21-hydroxylase. *Immunol. Rev.* 87: 39-60, 1985a.

Carroll, M.C., Palsdottir, A., Belt, K.T., Porter, R.R. Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. *EMBO J.* 4: 2547-2552, 1985b.

Carson, D.A. Rheumatoid factor. In: *Textbook of Rheumatology*, Kelley W.N., Harris, E.D., Ruddy, S., eds. W.B. Saunders, Philadelphia, pp. 677-690, 1981.

Christiansen, F.T., Dawkins, R.L., Uko, G., McCluskey, J., Kay, P.H., Zilko, P.J. Complement allotyping in SLE: association with C4A null. *Aust. NZ. J. Med.* 13: 483-488, 1983.

Cole, F.S., Whitehead, A.S., Auerbach, H.S., Lint, T., Seitz, H.J., Kilbridge, P., Colten, H.R. The molecular basis for genetic deficiency of the second component of human complement. *N. Engl. J. Med.* 313: 11-16, 1985.

Colten, H.R., Dowton, S.B. Regulation of complement gene expression. *Biochem. Soc. Symp.* 51: 37-46, 1986.

Curman, B., Ostberg, L., Sandberg, L., Malmheden-Ericksson, I., Stallenheim, G., Rask, L., Peterson, P.A. H-2 linked Ss protein is C4 component of complement. *Nature* 258: 243-245, 1975.

Dawkins, R.L., Christiansen, F.T., Kay, P.H., Garlepp, M., McCluskey, J., Hollingsworth, P.N., Zilko, P.J. Disease associations with complotypes, supratypes, and haplotypes. *Immunol. Rev.* 70: 522, 1983.

Dennis, P.M., Biegler, B., Papas, R. Improved measurement of monoclonal paraproteins in serum using agarose gel electrophoresis. *Ann. Clin. Biochem.* 24: 73-76, 1986.

Dodds, A.W., Law, S.K., Porter, R.R. The origin of the very variable haemolytic activities of the common human complement component C4 allotypes including C4A6. *EMBO J.* 4: 2239-2244, 1985.

Dubois, E.L., Wallace, D.J. *Dubois' lupus erythematosus.* Lea & Febiger, Philadelphia, pp. 161-271; 39-130, 1987.

Dumas, R., Hauptmann, G., Ohayon, E., Bascoul, S., Serre, A., Baldet, P., Naoumis, A. Deficit hereditaire du quatrieme composant du complement (C4), associe a un syndrome lupique. *Archives Francaises de Pediatrie.* 43: 267-269, 1986.

Dunham, I., Sargent, C.A., Trowsdale, J., Campbell, R.D. Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. *Proc. Natl. Acad. Sci. USA* 84: 7237-7241, 1987.

Dunham, I., Sargent, C.A., Dawkins, R.L., Campbell, R.D. Direct observation of the gene organisation of the complement C4 and 21-hydroxylase loci by pulsed-field gel electrophoresis. *J. Exp. Med.* (In press), 1989.

Dupont, B., Oberfield, S.E., Smithwick, E.M., Lee, T.D., Levine, L.S. Close genetic linkage between HLA and congenital adrenal hyperplasia (21-hydroxylase deficiency). *Lancet II*, 1309-1311, 1977.

Dyer, P., Grennan, D.M., Walton, K., Klimiuk, P., Dodds, W., Clague, R., Harris, R. HLA and properdin factor B (BF) in rheumatoid arthritis. *Ann. Rheum. Dis.* 43: 118, 1984.

Fielder, A.H.L., Walport, M.J., Batchelor, J.R., Rynes, F.I., Black, C.M., Dodi, I.A., Hughes, G.R.V. A family study of the major histocompatibility complex in patients with systemic lupus erythematosus. Importance of null alleles of C4A and C4B in determining disease susceptibility. *Br. Med. J.* 286: 425-428, 1983.

Fletcher, R.H., Fletcher, S.W., Wagner, E.H. *Clinical epidemiology: the essentials.* Williams & Wilkins, Baltimore. pp. 41-58, 1982.

Folk, J.E., Piez, K.A., Carroll, W.R., Gladner, J.A. Carboxypeptidase B: purification and characterization of the porcine enzyme. *J. Biol. Chem.* 235: 2272-2277, 1960.

Fu, S.M., Kunkel, H.G., Brusman, H.P., Allen, F.H. Jr., Fotino, M. Evidence for linkage between HL-A histocompatibility genes and those involved in the synthesis of the second component of complement. *J. Exp. Med.* 140: 1108-1110, 1974.

Giles, C.M., Gedde-Dahl, T., Jr., Robson, E.B., Thorsby, E., Olaisen, B., Arnason, A., Kissmeyer-Nielson, F., Schreuder, I. Rg (Rodgers) and the HL-A region; linkage and associations. *Tissue Antigens* 8: 143-149, 1976.

Giles, C.M. A new genetic variant for Chido. *Vox Sang.* 46: 149-152, 1984.

Goldstein, R., Arnett, F.C., McLean, R.H., Bias, W.B., Duvic, M. Molecular heterogeneity of complement component C4-null and 21-hydroxylase genes in systemic lupus erythematosus. *Arthritis Rheum.* 31: 736-744, 1988.

Gordon, J., Whitehead, H.R., Wormall, A. The action of ammonia on complement. The fourth component. *Biochem. J.* 20: 1028-1040, 1926.

Gregersen, P.K., Silver, J., Winchester, R.J. The shared epitope hypothesis - an approach to understanding the molecular genetics of rheumatoid arthritis susceptibility. *Arthritis Rheum.* 30: 1205-1213, 1987.

Gregersen, P.K., Shen, M., Song, Q.L., Merryman, P., Degar, S., Seki, T., Maccari, J., Goldberg, D., Murphy, H., Schwenzer, J., Wang, C.Y., Winchester, R.J., Nepom, G.T., Silver, J. Molecular diversity of HLA-DR4 haplotypes. *Proc. Natl. Acad. Sci. USA* 83: 2642-2646, 1986.

Hansen, T.H., Shreffler, D.C. Evidence for the involvement of the Ss protein of the mouse in the hemolytic complement system. *J. Exp. Med.* 141: 1216-1220, 1975.

Harrison, R.A., Thomas, M.L., Tack, B.F. Sequence determination of the thiolester site of the fourth component of human complement. *Proc. Natl. Acad. Sci. USA* 78: 7388-7392, 1981.

Hauptmann, G., Grooshans, E., Heid, E. Lupus erythemateux aigus et deficits hereditaires en complement. A propos d'un cas par deficit complet en C4. *Ann. Derm. Syphiligr.* 101: 479-496, 1974.

Hauptmann, G., Tappeiner, G., Schifferli, J.A. Inherited deficiency of the fourth component of human complement. *Immunodeficiency Reviews* 1: 3-22, 1988.

Hauptmann, G., Uring-Lambert, B., Vegnaduzzi-Lamouche, N., Mascart-Lemone, F. RFLP studies of 3 complete C4-deficient patients. *Complement* 4: 166, 1987.

Hauptmann, G., Uring-Lambert, B., Kazatchkine, M., Boigtlander, V., Bahmer, F., Seligmann, M. Decreased levels of complement components C4 and C2 in lupus erythematosus: genetic defect or complement activation? *Diagnostic Immunol.* 2: 63-64, 1984.

Hauptmann, G. Goetz, J., Uring-Lambert, B., Grosshans, E. Component deficiencies. 2. The fourth component. *Prog. Allergy* 39: 232-249, 1986.

Holme, E., Cross, S.J., Veitch, J., O'Neill, G.J., Whaley, K. Quantitation of human C4A and C4B, in serum and plasma by enzyme-linked immunoadsorbent assay. *Immunogenetics* 27: 295-297, 1988.

Howard, P.F., Hochberg, M.C., Bias, W.B., Arnett, F.C., McLean, R.H. Relationship between C4 null genes, HLA-D region antigens, and genetic susceptibilities to systemic lupus erythematosus in Caucasians and black Americans. *Am. J. Med.* 81: 187-193, 1986.

Isenman, D., Young, J.R. The molecular basis for the difference in immune hemolysis activity of the Chido and Rodgers isotypes of the human complement component C4. *J. Immunol.* 132: 3019-3027, 1984.

Kay, P.H., McCluskey, J., Christiansen, F.T., Feeney, D., McCann, V.J., Zilko, P.J., Dawkins, R.L., O'Neill, G.J. Complement allotyping reveals new genetic markers in rheumatoid arthritis and diabetes mellitus. *Tiss. Antigens* 21: 159-160, 1983.

Kemp, M.E., Atkinson, J.P., Skanes, V.M., Levine, R.P., Chaplin, D.D. Deletion of C4A genes in patients with systemic lupus erythematosus. *Arthritis Rheum.* 30: 1015-1022, 1987.

Kinoshita, T., Dodds, A.W., Law, S.K.A., Inoue, K. The basis of the low haemolytic activity of the C4A6 allotype. *Complement* 5: 182, 1988.

Kjellman, M., Laurell, A.B., Low, B., Sjöholm, A.G. Homozygous deficiency of C4 in a child with a lupus erythematosus syndrome. *Clin. Genet.* 22: 331-339, 1982.

Klein, G., Tappeiner, G., Hintner, H., Schultz, S., Wolff, K. Systemischer lupus erythematosus bei hereditarer defizienz der vierten komplement komponente. *Hautarzt* 35: 27-32, 1984.

Klitz, W., Borot, N., Thomson, G. A population analysis of C4 typing. *Hum. Immunol.* 23: 114, 1988.

Laurent, M.R., Welsh, K.I. Genetic markers in rheumatological diseases. *J. Immunogenet.* 10: 275-291, 1983.

Law, S.K.A., Dodds, A.W., Porter, R.R. A comparison of the properties of two classes, C4A and C4B, of the human complement component C4. *EMBO J.* 3: 1819-1823, 1984.

Maillet, F., Goetz, J., Hauptmann, G., Clauvel, J.P., Mery, J.Ph., Kahn, M.F., Kazatchkine, M. Les composants de la voie classique du complement au cours du lupus erythemateux dissemine. *Presse Medicale.* 16: 378-382, 1987.

Mancini, G., Carbonara, A.O., Heremans, J.F. Immunochemical quantitation of antigens by single radial immunodiffusion. *Immunochemistry* 2: 235-254, 1965.

Mascart-Lemone, F., Hauptmann, G., Goetz, J., Duchateau, J., Delespesse, G., Vray, B., Dab, I. Genetic deficiency of C4 presenting with recurrent infections and a SLE-like disease. Genetic and immunologic studies. *Am. J. Med.* 75: 295-304, 1983.

Mauff, G., Alper, C.A., Awdeh, Z., Batchelor, J.R., Bertrams, J., Bruun-Peterson, B., Dawkins, R.L., Demant, P., Edwards, J., Gross-Wilde, H., Hauptmann, G., Klouda, P., Lamm, L., Mollenhauer, E., Nerl, C., Olaisen, B., O'Neill, G., Rittner, C., Roos, M.H., Skanes, V., Teisberg, P., Wells, L. Statement on the nomenclature of human C4 allotypes. *Immunobiol.* 164: 184-191, 1983.

Mauff, G., Bender, K., Fischer, B. Genetic polymorphism of the fourth component of human complement. *Vox Sang.* 34: 296-301, 1978.

Meo, T., Krasteff, T., Shreffler, D.C. Immunochemical characterization of murine H-2 controlled Ss (serum substance) protein through identification of its human homologue as the fourth component of complement. *Proc. Natl. Acad. Sci. USA* 72: 4536-4540, 1975.

Meo, T., Atkinson, J.P., Bernoco, M., Bernoco, D., Ceppelini, R. Structural heterogeneity of C2 complement protein and its genetic variants in man: a new polymorphism of the HLA region. *Proc. Natl. Acad. Sci. USA* 74: 1672-1675, 1977.

Middleton, J., Crookston, M.C., Falk, J.A., Robson, E.B., Cook, P.J.L., Batchelor, J.R., Bodmer, J., Ferrara, G.B., Festenstein, H., Harris, R., Kissmeyer-Nielson, F., Lawler, S.D., Sachs, J.A., Wolf, E. Linkage of Chido and HL-A. *Tissue Antigens* 4: 366-373, 1974.

Minta, J.O., Urowitz, M.B., Gladman, D.D., Irizawa, T., Biggar, W.D. Selective deficiency of the fourth component of complement in a patient with systemic lupus erythematosus (SLE). Immunochemical and biological studies. *Clin. Exp. Immunol.* 45: 72-80, 1981.

Muir, W.A., Hedrick, S., Alper, C.A., Ratnoff, O.D., Schacter, B., Wisnieski, J.J. Inherited incomplete deficiency of the fourth component of complement (C4) determined by a gene not linked to human histocompatibility leukocyte antigens. *J. Clin. Invest.* 74: 1509-1514, 1984.

Naama, J.K., Mitchell, W.S., Whaley, K.S. Inhibition of complement-mediated solubilisation of antigen-antibody complexes by serum from patients with rheumatoid arthritis. *Clin. Exp. Immunol.* 54: 429-438, 1983.

Nagura, H., Hasegawa, H., Yoshimura, S., Watanabe, K. The third (C3) and fourth (C4) components of complement in human liver: immunocytochemical evidence for hepatocytes as the site of synthesis. *Acta. Pathol. Jpn.* 35: 71-78, 1985.

National Bureau of Standards, U.S. Department of Commerce. *Experimental Statistics. Handbook 91.* Chapter 17, pp. 1-5, 1966.

Nepom, G.T., Seyfried, C.E., Holbeck, S.L. Identification of HLA-Dw14 genes in DR4+ rheumatoid arthritis. *Lancet* II: 1002-1004, 1986.

O'Neill, G.J., Yang, S.Y., Dupont, B. Two HLA-linked loci controlling the fourth component of human complement. *Proc. Natl. Acad. Sci. USA* 75: 5165-5169, 1978a.

O'Neill, G.J., Yang, S.Y., Tegoli, J., Berger, R., Dupont, B. Chido and Rodgers blood groups are distinct antigenic components of human complement C4. *Nature* 273: 668-670, 1978b.

O'Neill, G.J., Nerl, C.W., Kay, P.H., Christiansen, F.T., McCluskey, J., Dawkins, R.L. Complement C4 is a marker for adult rheumatoid arthritis. *Lancet* II: 214, 1982.

Olaisen, B., Teisberg, P., Jonassen, R. The C4 system: quantitative studies of different genotypes. *Immunobiol.* 158: 82-85, 1980.

Palsdottir, A., Fossdal, R., Arnason, A., Edwards, H.H., Jensson, O. Heterogeneity of human C4 gene size. *Immunogenetics* 25: 299-304, 1987a.

Palsdottir, A., Arnason, A., Fossdal, R., Jensson, O. Gene organization of haplotypes expressing two different C4A allotypes. *Hum. Genet.* 76: 220-224, 1987b.

Partanen, J., Koskimies, S. Human MHC class III genes, Bf and C4. Polymorphism, complotypes and association with MHC class I genes in the Finnish population. *Human Hered.* 36: 269-275, 1986.

Partanen, J., van Assendelft, A.H.W., Koskimies, S., Forsberg, S., Hakala, M., Ilonen, J. Patients with rheumatoid arthritis and gold-induced pneumonitis express two high-risk major histocompatibility complex patterns. *Chest* 92: 277-281, 1987.

Pinals, R. Felty's syndrome. In: *Textbook of Rheumatology.* Kelley, W.N., Harris, E.D., Jr., Ruddy, S., Sledge, C.B., eds. W.B. Saunders, Philadelphia, pp. 964-968, 1981.

Porter, R.R. Complement polymorphism, the major histocompatibility complex and associated diseases: a speculation. *Mol. Biol. Med.* 1: 161-168, 1983.

Raum, D., Awdeh, Z., Anderson, J., Strong, L., Granados, J., Pevan, L., Giblett, E., Yunis, E.J., Alper, C.A. Human C4 haplotypes with duplicated C4A or C4B. *Am. J. Hum. Genet.* 36: 72-79, 1984.

Reid, K.B.M., Porter, R.R. The proteolytic activation systems of complement. *Annu. Rev. Biochem.* 50: 433, 1981.

Reveille, J.D., Arnett, F.C., Wilson, R.W., Bias, W.B., McLean, R.H. Null alleles of the fourth component of complement and HLA haplotypes in familial systemic lupus erythematosus. *Immunogenetics* 21: 299-311, 1985.

Rittner, C., Hauptmann, G., Gross-Wilde, H., Grosshans, E., Tongio, M.M., Mayer, S. Linkage between HL-A (major histocompatibility complex) and genes controlling synthesis of the fourth component of complement. In: *Histocompatibility testing 1975.* Kissmeyer-Nielsen, F., ed. Munksgaard, Copenhagen pp. 945-953, 1975.

Rittner, C., Mauff, G. C4 polymorphism. In: *Histocompatibility testing 1984.* Albert, E.D., Baur, M.P., Mayr, W.R., eds. Springer-Verlag, pp. 318-324, 1984a.

Rittner, C., Giles, C.M., Roos, M.H., Demant, P., Mollenhauer, E. Genetics of human C4 polymorphism: detection and segregation of rare and duplicated haplotypes. *Immunogenetics* 19: 321-333, 1984b.

Robinson, M.A., Carroll, M.C., Johnson, A.H., Hartzman, R.J., Belt, K.T., Kindt, T.J. Localization of C4 genes within the HLA complex by molecular genotyping. *Immunogenetics* 21: 143-152, 1985.

Rosenfeld, S.I., Ruddy, S., Austin, K.F. Structural polymorphism of the fourth component of human complement. *J. Clin. Invest.* 48: 2283-2292, 1969.

Ross, G.D. In: *Immunobiology of the complement system. An introduction for research and clinical medicine.* Ross, G.D. ed. Academic Press Inc., New York, 1986.

Sacha, J.A., Kirwan, J.R. Multiple HLA associations and disease susceptibility. *Dis. Markers.* 4: 13-17, 1986.

Schaller, J.G., Gilliland, B.G., Ochs, H.D., Leddy, J.P., Agodoa, L.C.Y., Rosenfeld, S.I. Severe systemic lupus erythematosus with nephritis in a boy with deficiency of the fourth component of complement. *Arthritis Rheum.* 20: 1519-1525, 1977.

Schendel, D.J., O'Neill, G., Wank, R. MHC-linked class III genes. Analysis of C4 gene frequencies, complotypes and associations with distinct HLA haplotypes in German Caucasians. *Immunogenetics* 20: 23-31, 1984.

Schifferli, J.A., Steiger, G., Paccaud, J.P., Sjöholm, A.G. Difference in the biological properties of the two forms of the fourth component of human complement (C4). *Clin. Exp. Immunol.* 63: 473-477, 1986.

Schneider, P.M., Carroll, M.C., Alper, C.A., Rittner, C., Whitehead, A.S., Yunis, E.J., Colten, H.R. Polymorphism of human complement C4 and steroid 21-hydroxylase gene. Restriction fragment length polymorphisms revealing structural deletions, homoduplications and size variants. *J. Clin. Invest.* 78: 650-657, 1986.

Siegel, S. *Non-parametric statistics for the behavioural sciences. McGraw-Hill series in psychology.* McGraw-Hill Book Company, New York, pp. 184-194, 1956.

Shreffler, D.C., Owen, R.D. A serologically detected variant in mouse serum: inheritance and association with the histocompatibility-2 locus. *Genetics* 48: 9-25, 1963.

Shreffler, D.C., Atkinson, J.P., Chan, A.C., Karp, D.R., Killion, C.C., Ogata, R.T., Rosa, P.A. The C4 genes and the SLP-genes of the complement region of the murine H-2 major histocompatibility complex. *Philos. Trans. R. Soc. London B.* 306-395-403, 1984.

Sim, R.B., Sim, E. Autolytic fragmentation of complement components C3 and C4 under denaturing conditions, a property shared with alpha-2-macroglobulin. *Biochem. J.* 193: 129-141, 1981.

Sim, E., Cross, S.J. Phenotyping of human complement component C4, a class-III HLA antigen. *Biochem. J.* 239: 763-767, 1986.

Sjöholm, A.G., Kjellman, N.-I.M., Low, B. C4 allotypes and HLA-DR antigens in the family of a patient with C4 deficiency. *Clin. Genet.* 28: 385-393, 1985.

Stastny, P. Mixed lymphocyte culture in rheumatoid arthritis. *J. Clin. Invest.* 57: 1148-1164, 1976.

Stastny, P. Association of the B-cell alloantigen in DRw4 with rheumatoid arthritis. *N. Engl. J. Med.* 298: 869-873, 1978.

Suzuki, D.T., Griffiths, A.J.F., Miller, J.H., Lewontin, R.C. *An introduction to genetic analysis.* W.H. Freeman and Company, New York, pp. 545-547, 1986.

Takahashi, M., Takahashi, S., Brade, V., Nussenzweig, V. Requirements for solubilization of immune aggregates by complement: the role of the classical pathway. *J. Clin. Invest.* 62: 349-358, 1978.

Tappeiner, G. Schloz, S., Linert, J., Albert, E., Wolff, K. Hereditary deficiency of the fourth component of complement (C4): study of a family. Cutaneous immunopathology. *Inserm* 800: 399-404, 1978.

Tappeiner, G., Hintner, H., Scholtz, S., Albert, E., Linert, J., Wolff, K. Systemic lupus erythematosus in hereditary deficiency of the fourth component of complement. *J. Am. Acad. Derm.* 7: 66-79, 1982.

Teisberg, P.I., Akesson, B., Olaisen, T., Gedde-Dahl, T. Jr., Thorsby, E. Genetic polymorphism of C4 in man and localization of a structural C4 locus to the HLA gene complex on chromosome 6. *Nature* 264: 253-254, 1976.

Teisberg, P., Olaisen, B., Jonassen, R., Gedde-Dahl, T., Jr., Thorsby, E. The genetic polymorphism of the fourth component of human complement: methodological aspects and a presentation of linkage and association data relevant to its localization in the HLA region. *J. Exp. Med.* 146: 1380-1389, 1977.

Thomson, W., Sanders, P.A., Martin, D., Davidson, J., Dyer, P.A., Grennan, D.M. Complement C4B-null alleles in Felty's syndrome. *Arthritis Rheum.* 31: 984-989, 1988.

Thorndike, R.M. Discriminant analysis. In: *Correlational procedures for research.* Gardner Press Inc. New York, pp. 203-220, 1978.

Truedsson, L., Avedeh, Z., Yunis, E.J., Mrose, S., Alper, C.A. Most extended MHC haplotypes have duplications and/or deletions of C4 genes. *Complement* 4: 232, 1987.

Uko, G., Christiansen, F.T., Dawkins, R.L., McCann, V.J. Reference ranges for serum C4 concentrations in subjects with and without C4 null alleles. *J. Clin. Pathol.* 39: 573-576, 1986.

Uring-Lambert, B., Vegnaduzzi, N., Carroll, M.C., Tongio, M.M., Goetz, J., Hauptmann, G. Heterogeneity in the structural basis of the human complement C4A null allele (C4A*Q0) as revealed by Hind III restriction fragment length polymorphism analysis. *FEBS Letters* 217: 65-68, 1987.

Uring-Lambert, B., Goetz, J., Tongio, M.M., Mayer, S., Hauptmann, G. C4 haplotypes with duplication at the C4A or C4B loci: frequency and associations with Bf, C2 and HLA-A, B, C, DR alleles. *Tiss. Antigens* 24: 70-72, 1984.

Weitkamp, L.R., Lamm, L.U. Report of the committee on the genetic constitution of chromosome 6. *Cytogenet. Cell Genet.* 32: 130-143, 1982.

Welch, T.R., Beischel, L., Berry, A., Forristal, J., West, C.D. The effect of null C4 alleles on complement function. *Clin. Imm. and Immunopath.* 34: 316-325, 1985.

White, P.C., Grossberger, D., Onufer, B.J., Chaplin, D.D., New, M.I., Dupont, B., Strominger, J.L. Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man. *Proc. Natl. Acad. Sci. USA* 82: 1089-1093, 1985.

Whitehead, A.S., Goldberger, G., Woods, D.E., Markham, A.F., Colten, H.R. Use of a cDNA clone for the fourth component of human complent (C4) for analysis of a genetic deficiency of C4 in guinea pig. *Proc. Natl. Acad. Sci. USA* 80: 5387-5391, 1983.

Wisnieski, J.J., Nathanson, M.H., Anderson, J.E., Davis, A.E., Alper, C.A., Naff, G.B. Metabolism of C4 and linkage analysis in a kindred with inherited incomplete C4 deficiency. *Arthritis Rheum.* 30: 919-926, 1987.

Yu, Y.C., Campbell, R.D., Porter, R.R. A structural model for the location of the Rodgers and the Chido antigenic determinants and their correlation with the human complement component C4A/C4B isotypes. *Immunogenetics* 27: 399-405, 1988.

Yu, C.Y., Belt, K.T., Giles, C.M., Campbell, R.D., Porter, R.R. Structural basis of the polymorphism of human complement C4A and C4B: gene size, reactivity and antigenicity. *EMBO J.* 5: 2873-2881, 1986.

Yu, Y.C., Campbell, R.D. Definitive RFLP's to distinguish between the human complement C4A/C4B isotypes and the major Rodgers/Chido determinants: application to the study of C4 null alleles. *Immunogenetics* 25: 383-390, 1987.

Zhang, W.J., Kay, P.H., Cobain, T.J., Dawkins, R.L. C4 allotyping on plasma or serum: application to routine laboratories. *Hum. Immunol.* 21: 165-171, 1988.

Appendix A

SYSTAT output from discriminant analysis of C4A:C4B ratios for genotyped panel subjects

	Factor	Distance(1)	Distance(2)	Distance(3)	Prob(1)
	Prob(2)	Prob(3)	Group	Predict	Group$
1	2.274	3.375	2.303	0.244	0.003
1	0.067	0.929	2.000	3.000	
2	3.276	4.376	3.305	0.757	0.000
2	0.006	0.994	2.000	3.000	
3	-0.123	0.978	0.094	2.642	0.377
3	0.605	0.019	2.000	2.000	
4	0.017	1.118	0.046	2.501	0.339
4	0.633	0.028	2.000	2.000	
5	0.569	1.669	0.598	1.950	0.201
5	0.678	0.121	2.000	2.000	
6	-0.160	0.941	0.131	2.678	0.387
6	0.597	0.017	2.000	2.000	
7	-0.249	0.851	0.220	2.768	0.411
7	0.576	0.013	2.000	2.000	
8	0.282	1.383	0.311	2.237	0.271
8	0.671	0.058	2.000	2.000	
9	-1.459	0.358	1.429	3.977	0.722
9	0.277	0.000	1.000	1.000	
10	-0.819	0.282	0.790	3.337	0.566
10	0.431	0.002	2.000	1.000	
11	-0.427	0.674	0.397	2.945	0.460
11	0.533	0.008	2.000	2.000	
12	-0.573	0.528	0.544	3.092	0.500
12	0.495	0.005	2.000	1.000	
13	-0.129	0.972	0.100	2.648	0.378
13	0.604	0.018	2.000	2.000	
14	0.455	1.556	0.484	2.064	0.228
14	0.681	0.091	2.000	2.000	
15	0.895	1.996	0.924	1.624	0.129
15	0.618	0.253	2.000	2.000	
16	-0.439	0.661	0.410	2.958	0.463
16	0.530	0.007	2.000	2.000	
17	-0.468	0.633	0.439	2.987	0.471
17	0.522	0.007	2.000	2.000	
18	0.223	1.324	0.252	2.296	0.286
18	0.665	0.049	2.000	2.000	
19	0.863	1.964	0.892	1.656	0.136
19	0.627	0.237	2.000	2.000	
20	0.643	1.744	0.672	1.876	0.184
20	0.671	0.145	2.000	2.000	
21	-0.459	0.642	0.430	2.978	0.468

21	0.525	0.007	2.000	2.000	
22	-1.004	0.097	0.975	3.523	0.615
22	0.384	0.001	2.000	1.000	
23	-0.369	0.732	0.340	2.888	0.444
23	0.547	0.009	2.000	2.000	
24	-1.486	0.385	1.457	4.005	0.728
24	0.271	0.000	1.000	1.000	
25	0.002	1.103	0.031	2.517	0.343
25	0.630	0.027	1.000	2.000	
26	-1.448	0.347	1.419	3.967	0.720
26	0.280	0.000	1.000	1.000	
27	-1.283	0.182	1.254	3.802	0.683
27	0.316	0.001	1.000	1.000	
28	-1.317	0.216	1.288	3.836	0.691
28	0.309	0.000	1.000	1.000	
29	-1.142	0.042	1.113	3.661	0.649
29	0.350	0.001	1.000	1.000	
30	-1.391	0.290	1.362	3.909	0.708
30	0.292	0.000	1.000	1.000	
31	-1.181	0.080	1.152	3.700	0.659
31	0.340	0.001	1.000	1.000	
32	-0.305	0.796	0.276	2.824	0.426
32	0.563	0.011	1.000	2.000	
33	-1.491	0.390	1.462	4.010	0.729
33	0.270	0.000	1.000	1.000	
34	-1.469	0.368	1.440	3.988	0.725
34	0.275	0.000	1.000	1.000	
35	-1.266	0.166	1.237	3.785	0.679
35	0.320	0.001	1.000	1.000	
36	-1.536	0.435	1.507	4.055	0.739
36	0.261	0.000	1.000	1.000	
37	-1.333	0.232	1.304	3.852	0.695
37	0.305	0.000	1.000	1.000	
38	-0.853	0.247	0.824	3.372	0.576
38	0.422	0.002	1.000	1.000	
39	-1.483	0.382	1.454	4.002	0.728
39	0.272	0.000	1.000	1.000	
40	-0.905	0.196	0.876	3.423	0.589
40	0.409	0.002	1.000	1.000	
41	-1.142	0.041	1.113	3.661	0.649
41	0.350	0.001	1.000	1.000	
42	-1.318	0.218	1.289	3.837	0.691
42	0.308	0.000	1.000	1.000	
43	-1.162	0.061	1.133	3.681	0.654
43	0.345	0.001	1.000	1.000	
44	-1.446	0.346	1.417	3.965	0.720
44	0.280	0.000	1.000	1.000	
45	-1.459	0.358	1.430	3.978	0.723
45	0.277	0.000	1.000	1.000	
46	-1.250	0.149	1.221	3.769	0.675
46	0.324	0.001	1.000	1.000	
47	-1.250	0.149	1.221	3.769	0.675
47	0.324	0.001	1.000	1.000	
48	-0.981	0.119	0.952	3.500	0.609
48	0.390	0.001	1.000	1.000	
49	-1.102	0.001	1.073	3.620	0.639

49	0.360	0.001	1.000	1.000	
50	-1.341	0.240	1.312	3.860	0.696
50	0.303	0.000	1.000	1.000	
51	-1.289	0.188	1.260	3.808	0.685
51	0.315	0.000	1.000	1.000	
52	4.514	5.615	4.543	1.995	0.000
52	0.000	1.000	3.000	3.000	
53	1.098	2.199	1.127	1.421	0.091
53	0.539	0.370	3.000	2.000	
54	2.213	3.313	2.242	0.306	0.004
54	0.078	0.918	3.000	3.000	
55	4.549	5.649	4.578	2.030	0.000
55	0.000	1.000	3.000	3.000	
56	0.500	1.600	0.529	2.019	0.217
56	0.681	0.102	3.000	2.000	
57	2.672	3.773	2.701	0.154	0.001
57	0.026	0.974	3.000	3.000	
58	2.541	3.642	2.570	0.023	0.001
58	0.035	0.963	3.000	3.000	
59	2.648	3.748	2.677	0.129	0.001
59	0.027	0.972	3.000	3.000	
60	2.247	3.348	2.276	0.272	0.004
60	0.072	0.925	3.000	3.000	
61	4.371	5.471	4.400	1.852	0.000
61	0.000	1.000	3.000	3.000	
62	5.799	6.900	5.828	3.280	0.000
62	0.000	1.000	3.000	3.000	
63	2.663	3.763	2.692	0.144	0.001
63	0.026	0.973	3.000	3.000	
64	0.324	1.425	0.353	2.195	0.260
64	0.675	0.065	3.000	2.000	
65	4.430	5.531	4.459	1.911	0.000
65	0.000	1.000	3.000	3.000	
66	4.618	5.719	4.647	2.099	0.000
66	0.000	1.000	3.000	3.000	
67	-1.728	0.628	1.699	4.247	0.777
67	0.223	0.000	1.000	1.000	
68	-1.129	0.028	1.100	3.648	0.646
68	0.353	0.001	1.000	1.000	
69	-1.282	0.181	1.253	3.801	0.683
69	0.317	0.001	1.000	1.000	
70	-0.822	0.278	0.793	3.341	0.567
70	0.431	0.002	1.000	1.000	
71	-0.859	0.241	0.830	3.378	0.577
71	0.421	0.002	1.000	1.000	
72	-1.482	0.382	1.453	4.001	0.728
72	0.272	0.000	1.000	1.000	
73	-1.910	0.809	1.881	4.429	0.809
73	0.191	0.000	1.000	1.000	
74	0.114	1.215	0.143	2.405	0.314
74	0.650	0.036	1.000	2.000	
75	-1.289	0.188	1.260	3.808	0.685
75	0.315	0.000	1.000	1.000	
76	-1.273	0.172	1.244	3.792	0.681
76	0.319	0.001	1.000	1.000	
77	0.374	1.474	0.403	2.145	0.248

77	0.678	0.074	1.000	2.000	
78	-0.571	0.530	0.542	3.090	0.499
78	0.496	0.005	1.000	1.000	
79	0.049	1.150	0.078	2.470	0.331
79	0.639	0.030	1.000	2.000	
80	-0.973	0.128	0.944	3.492	0.607
80	0.392	0.001	1.000	1.000	
81	-1.329	0.228	1.300	3.848	0.694
81	0.306	0.000	1.000	1.000	
82	0.006	1.106	0.035	2.513	0.342
82	0.631	0.027	1.000	2.000	
83	-0.998	0.103	0.969	3.517	0.613
83	0.385	0.001	1.000	1.000	
84	-1.544	0.444	1.515	4.063	0.741
84	0.259	0.000	1.000	1.000	
85	-0.290	0.810	0.261	2.809	0.422
85	0.567	0.011	2.000	2.000	
86	-0.574	0.526	0.545	3.093	0.500
86	0.495	0.005	2.000	1.000	
87	0.351	1.452	0.380	2.168	0.254
87	0.677	0.069	2.000	2.000	
88	-0.776	0.325	0.747	3.294	0.555
88	0.443	0.003	2.000	1.000	
89	-0.419	0.682	0.390	2.938	0.457
89	0.535	0.008	2.000	2.000	
90	-0.602	0.499	0.573	3.120	0.508
90	0.488	0.004	2.000	1.000	
91	-0.196	0.905	0.167	2.715	0.396
91	0.589	0.015	2.000	2.000	
92	0.594	1.694	0.623	1.925	0.195
92	0.676	0.129	2.000	2.000	
93	-0.245	0.856	0.216	2.764	0.410
93	0.577	0.013	2.000	2.000	
94	-1.227	0.126	1.198	3.746	0.670
94	0.330	0.001	2.000	1.000	
95	-0.285	0.816	0.256	2.804	0.421
95	0.568	0.012	2.000	2.000	
96	-0.448	0.653	0.419	2.967	0.465
96	0.528	0.007	2.000	2.000	
97	0.868	1.969	0.897	1.651	0.135
97	0.626	0.239	2.000	2.000	
98	-0.617	0.483	0.588	3.136	0.512
98	0.484	0.004	2.000	1.000	
99	-0.651	0.450	0.622	3.170	0.521
99	0.475	0.004	2.000	1.000	
100	-0.634	0.467	0.605	3.153	0.516
100	0.480	0.004	2.000	1.000	
101	0.124	1.225	0.153	2.395	0.311
101	0.651	0.037	2.000	2.000	
102	-0.415	0.686	0.386	2.934	0.456
102	0.536	0.008	2.000	2.000	
103	3.741	4.842	3.771	1.223	0.000
103	0.002	0.998	3.000	3.000	
104	0.831	1.932	0.860	1.688	0.143
104	0.636	0.221	3.000	2.000	
105	2.415	3.516	2.444	0.104	0.002

105	0.048	0.950	3.000	3.000	
106	1.407	2.508	1.436	1.112	0.046
106	0.380	0.574	3.000	3.000	
107	-1.361	0.261	1.332	3.880	0.701
107	0.299	0.000	3.000	1.000	
108	0.678	1.778	0.707	1.841	0.176
108	0.667	0.157	3.000	2.000	

Appendix B

Haplotype frequencies* for phenotyped population members

Population	n	fAXBX	fAQ0BX	fAXBQ0
Newfoundland controls	120	0.6593	0.1826	0.1581
Newfoundland RA	92	0.5115	0.2331	0.2554
St. Louis controls	398	0.6832	0.1227	0.1941
St. Louis SLE	96	0.5333	0.3225	0.1442
Hungarian controls	38	0.4898	0.2809	0.2293
Hungarian SLE	42	0.7366	0.1543	0.1091

* Haplotype frequencies were calculated according to the observed frequency of homozygous null C4A and homozygous null C4B in each population (pp. 46, 88; Table 4.23). A detailed example showing the calculation of haplotype and genotype frequencies can be found on pp. 93-95.